78-39

Sgríbhinní Institiúid Árd-Léinn Bhaile Átha Cliath
Sraith. A, Uimh. 25

Communications of the Dublin Institute for Advanced Studies
Series A (Theoretical Physics), No. 25

DIFFERENTIAL GEOMETRY OF INSTANTONS

By

J. H. Rawnsley

DUBLIN

Institiúid Árd-Léinn Bhaile Átha Cliath
Dublin Institute for Advanced Studies
1978

Price: £3.15

Preface

In Spring 1977 I gave a series of lectures at DIAS on the new mathematical results in the theory of instantons. These lectures covered the calculation by Atiyah, Hitchin, and Singer of the dimension of the space of self-dual instantons by means of the Index Theorem applied to a suitable complex. This Communication is an expanded version of the notes for those lectures.

I have added new sections on related topics, which include holonomy groups, characteristic classes, and the explicit construction of self-dual instantons by Atiyah, Hitchin, Drinfeld, and Manin; the last chapter is based on lectures I gave at Imperial College, London, in February 1978, and at the Mathematical Institute, Oxford, in May 1978.

Table of Contents

0. Introduction

0.1. Finite-action solutions of the Yang-Mills field equations,

$$\sum_\mu D_\mu F_{\mu\nu} = 0, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu], \tag{0.1}$$

have been the object of intense investigation by both mathematicians and physicists since the first solution was found by Belavin, Polyakov, Schwartz, Tyupkin (1976). In physics such solutions are critical points for the action of a gauge field, and so their presence may have a contribution to Feynman path integrals over such fields. Mathematically, the field equations (0.1) are a system of non-linear partial differential equations, and it is relatively rarely that such equations can be solved explicitly. A study of the method of solution may throw some light on the theory of non-linear equations in general.

This Communication is intended as an introduction to some of the branches of differential geometry and global analysis which have recently been applied to the study of the Yang-Mills equations. We shall not cover any of the physical background. Interested readers should consult the survey article of Coleman (1978).

After the original lectures were given much more progress was made, including the transcription by Atiyah & Ward (1977) of the self-dual solutions into algebraic bundles, and the construction of these solutions (Atiyah, Hitchin, Drinfeld, Manin, 1978) using results of Horrocks (1964) and Barth & Hulek (1978). An extra chapter (Chapter 5) which briefly describes these results has therefore been added.

As this Communication is intended to be introductory, the more elementary theory is described in some detail, with advanced results only sketched, or given as references. The reader is encouraged to consult the various texts cited for further information.

0.2. Let $\mathfrak{g}$ be the Lie algebra of a compact Lie group G. A Yang-Mills potential (Yang & Mills, 1954) is a smooth 4-vector of $\mathfrak{g}$ -valued fields $A_\mu(x)$ on $\mathbb{R}^4$. The associated action S(A) is given by

$$S(A) = \tfrac{1}{2}\int_{\mathbb{R}^4} \sum_{\mu\nu} \mathrm{Tr}[F_{\mu\nu} F^{\mu\nu}]\, d^4x ,$$

where for simplicity we suppose G is a matrix group. Here

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \tag{0.2}$$

are the field strengths. If we replace A_μ by A'_μ , where

$$A'_\mu = g^{-1}A_\mu g + g^{-1}\partial_\mu g \tag{0.3}$$

with $g : \mathbb{R}^4 \to G$, then

$$F'_{\mu\nu} = g^{-1}F_{\mu\nu}g$$

and S(A') = S(A). These are gauge transformations, and the rule (0.3) can be interpreted as saying that A_μ is a connection form in some principal G-bundle. $F_{\mu\nu}$ is the curvature form of this connection.

Varying A_μ we obtain the field equations

$$\frac{\delta S}{\delta A_\nu(x)} = \sum_\mu \partial_\mu F_{\mu\nu} + [A_\mu , F_{\mu\nu}] = 0. \tag{0.4}$$

If we define an operator

$$D_\mu = \partial_\mu + [A_\mu , \cdot] \tag{0.5}$$

then (0.4) may be written

$$\sum_\mu D_\mu F_{\mu\nu} = 0.$$

(0.5) is the covariant derivative defined by A_μ, and $F_{\mu\nu}$ satisfies Bianchi's identity

$$\sum_\mu D_\mu(*F)_{\mu\nu} = 0 \tag{0.6}$$

where

$$(*F)_{\mu\nu} = \tfrac{1}{2}\sum_{\rho\sigma} \varepsilon_{\mu\nu\rho\sigma}F^{\rho\sigma} .$$

It follows that, if F = *F, then the field equations (0.4) are an automatic consequence of Bianchi's identity. These fields are called self-dual and are the principal objects of our interest.

0.3. At this point we should mention that we are considering the Yang-Mills equations in Euclidean rather than Minkowski space, because in Euclidean space they are rather more regular (elliptic instead of hyperbolic, at least in a suitable gauge). In this setting they may be relevant to the analytic continuation of the Green's functions of the field theory to the Euclidean region. Also, in order to avoid growth conditions at infinity to make S(A) convergent, we shall work on the conformal compactification S^4 of $\mathbb{R}^4$. As the equations (0.6) are conformally invariant, any solution of (0.6) on S^4 gives a solution on $\mathbb{R}^4$ with $S(A) < \infty$. It is not known if the converse is true.

On S^4, global topological properties come into play. There are non-trivial principal bundles over S^4, and if we assume for simplicity that $G = SU(2)$, these are classified by the integers $\mathbb{Z}$. The integer corresponding with a given bundle may be determined from the Yang-Mills field:

$$k = \frac{1}{16\pi^2} \int \sum_{\mu\nu} \mathrm{Tr}[F_{\mu\nu}(*F)^{\mu\nu}]\, d^4x.$$

k may be interpreted as the Chern class of the vector bundle associated to the standard representation of $SU(2)$ on $\mathbb{C}^2$. See the text and the appendix for details. Care should be taken over normalizations, since various other numerical factors appear in the literature, in front of the integral, depending on how $F_{\mu\nu}$ is related to its coordinate invariant definition F.

For any connection A one always has

$$S(A) \geq 8\pi^2 |k|,$$

and equality holds (for $k > 0$) if and only if $F = *F$. Thus the self-dual solutions represent the absolute minimum of $S(A)$ over all connections which may be defined in a fixed principal bundle (in physical terminology, all Yang-Mills potentials with a given topological charge).

The solution of Belavin et al. (1976) had charge $k = 1$, and 't Hooft (unpublished) generalized their construction to obtain self-dual solutions for all $k \geq 1$ for $SU(2)$. For charge k he found a $5k$-parameter family which could not be obtained from each other by gauge transformations of the form (0.3) (we call these gauge-inequivalent). Jackiw, Nohl, Rebbi (1977) used conformal invariance to extend this to a $5k + 4$ parameter family for $k \geq 3$ (5 for $k = 1$, 13 for $k = 2$), and it was natural to ask if this represented the largest possible family of inequivalent solutions.

This question was studied by linearizing the equations $F = *F$ around a known solution. If the general connection A_μ is written $\mathring{A}_\mu + a_\mu$, with $\mathring{A}_\mu$ the known solution, then A_μ has self-dual curvature if and only if

$$(D_\mu a_\nu - D_\nu a_\mu + [a_\mu, a_\nu])_- = 0, \tag{0.7}$$

where

$$(f_{\mu\nu})_- = \tfrac{1}{2}[f_{\mu\nu} - \sum_{\rho\sigma} \tfrac{1}{2}\varepsilon_{\mu\nu\rho\sigma} f^{\rho\sigma}]$$

is the anti-self-dual part of a 2-form $f_{\mu\nu}$.

If the gauge is fixed covariantly by requiring

$$\sum_\mu D_\mu a_\mu = 0, \tag{0.8}$$

and (0.7) is linearized to give

$$(D_\mu a_\nu - D_\nu a_\mu)_- = 0, \tag{0.9}$$

then it has been shown by Atiyah, Hitchin & Singer (1977a,b), Schwarz (1977), and Jackiw & Rebbi (1977) that the system given by (0.8) and (0.9) is elliptic and the Index Theorem gives $8k - 3$ as the dimension of the space of solutions.

By using deformation theory techniques of Kuranishi (1965), Atiyah et al. (1977b) showed that these infinitesimal solutions are tangent to actual solutions, and that the space of gauge inequivalent solutions of $F = *F$, for $SU(2)$ is a smooth manifold of dimension $8k - 3$. A similar result was established for every simple group under the assumption of irreducibility of the connection (otherwise there are singularities present) (Atiyah et al. (1977b), Bernard, Christ, Guth, Weinberg (1977)).

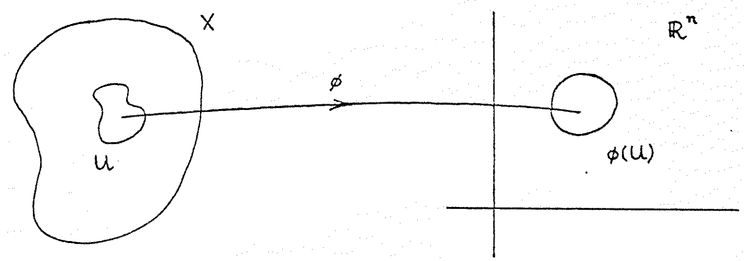
0.4. It can be seen that for $k = 1, 2$, $8k - 3$ agrees with the number of known solutions, 5 and 13, but disagrees for $k \geq 3$. This suggests two questions: Do the known solutions for $k = 1, 2$ account for all solutions for these values of k? How do we find the missing solutions for $k \geq 3$? It became possible to answer these questions when Atiyah & Ward (1977) translated the problem into one of algebraic geometry. Hartshorne (1978) showed that the known solutions for $k = 1$ and 2 represented all solutions for this charge, up to gauge equivalence. Then Atiyah et al. (1978) used results of Horrocks (1964) and Barth & Hulek (1978) to give an explicit solution for all self-dual Yang-Mills potentials in terms of linear algebra. One consequence of their result is that any self-dual connection is obtained by taking a suitable embedding in a trivial bundle and using the connection induced by the trivial connection.

0.5. The contents of the Communication may be summarized as follows: Chapter 1 deals with the elementary notions of a manifold, differential forms, exterior differentiation, metrics, orientation, duality, and various generalizations. Chapter 2 is a short description of vector bundles, covariant derivatives, differential operators, and the Index Theorem. In Chapter 3 we describe the theory of principal bundles, connections, and their holonomy groups. Chapter 4 covers the deformation theoretical calculations leading to the dimension 8k - 3 of the space of inequivalent self-dual SU(2) Yang-Mills fields. Chapter 5 summarizes the Atiyah & Ward (1977) transformation, and the explicit construction of Atiyah et al. (1978).

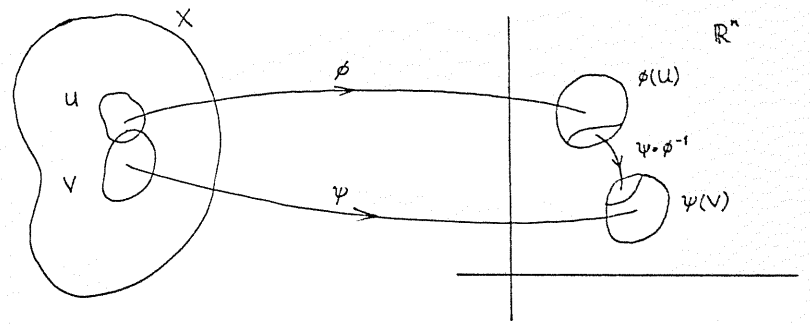
I would like to thank Professor J. T. Lewis for his advice and encouragement in producing this Communication. I would also like to thank my colleagues at the Dublin Institute for Advanced Studies, especially Professor L. O'Raifeartaigh, Dr. M. Scheunert and Dr. D. H. Tchrakian, for their interest in my lectures and for many useful conversations on Yang-Mills theory. I am grateful to Professor M. F. Atiyah and Dr. N. Hitchin for explaining their beautiful results to me. The manuscript was edited by Miss E. R. Wills and typed by Mrs. E. Maguire and I am very grateful to them for the speed and efficiency with which this was done.

1. DIFFERENTIAL CALCULUS ON MANIFOLDS

1.1 A manifold is a topological space consisting of pieces of Euclidean space 'glued' together by means of differentiable coordinate changes. The formal definition goes as follows: We say a map f between open subsets of two Euclidean spaces is *smooth* if all the partial derivatives of all orders of all the components of f exist, and are continuous. Let X be a topological space, and n an integer. An n-dimensional *chart* on X is a pair (U, ϕ) consisting of an open subset U of X and a homeomorphism ϕ of U onto an open subset of $\mathbb{R}^n$. U is called the domain of the chart (U, ϕ).



Two n-dimensional charts (U, ϕ), (V, ψ) on X are said to be *compatible* if $U \cap V = \emptyset$, or if $U \cap V \neq \emptyset$ and $\psi \circ \phi^{-1} : \phi(U \cap V) \to \psi(U \cap V)$ is smooth.



An n-dimensional *atlas* on X is a collection $\mathcal{A}$ of n-dimensional charts which are pairwise compatible, and whose domains cover X. For example, $\mathbb{R}^n$ itself has an atlas consisting of $U = \mathbb{R}^n$ and ϕ the identity map.

An atlas is maximal if it contains every chart compatible with all of its charts. Every atlas is contained in a unique maximal atlas. An n-dimensional *manifold* is a topological space X together with an n-dimensional maximal atlas on X. The atlas is usually fixed, and one refers to X itself as the manifold, but the same topological space can have more than one maximal atlas. However, the dimension depends only on the topology. It is usual to require X to be Hausdorff as a topological space. Note that, by the uniqueness of a maximal atlas containing a given atlas, it is sufficient to give one atlas on X to give a topological space X the structure of a manifold. Thus $\mathbb{R}^n$ becomes an n-dimensional manifold with respect to the maximal atlas containing the chart $(\mathbb{R}^n, \mathrm{id})$.

1.2. A slightly less trivial example of an atlas can be constructed as follows: Denote by S^n the set of unit vectors in $\mathbb{R}^{n+1}$. This is the n-dimensional sphere. If $x = (x^1, \ldots, x^{n+1}) \in \mathbb{R}^{n+1}$, then $x \in S^n$ if

$$x \cdot x = \sum_{i=1}^{n+1} (x^i)^2 = 1.$$

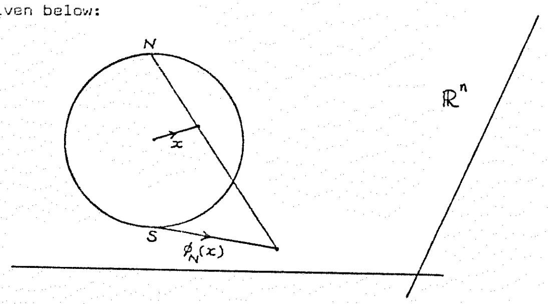
Let N denote the point $(0, \ldots, 0, 1)$ and S the point $(0, \ldots, 0, -1)$. Put $U_N = S^n - \{N\}$, $U_S = S^n - \{S\}$, and define maps

$$\phi_N : U_N \to \mathbb{R}^n, \quad \phi_S : U_S \to \mathbb{R}^n,$$

by

$$\phi_N(x) = \frac{2}{1-x^{n+1}}(x^1, \ldots, x^n), \quad \phi_S(x) = \frac{2}{1+x^{n+1}}(x^1, \ldots, x^n).$$

A picture of ϕ_N is given below:



Clearly U_N and U_S together cover S^n, whilst ϕ_N and ϕ_S map $U_N \cap U_S$ to the non-zero vectors in $\mathbb{R}^n$. We have

$$\phi_S \circ \phi_N^{-1}(y) = \frac{4y}{y \cdot y}, \quad y \in \mathbb{R}^n - \{0\},$$

which is smooth. Thus $\{(U_N, \phi_N), (U_S, \phi_S)\}$ is an n-dimensional atlas on S^n.

1.3. Let X be an n-dimensional manifold with atlas $\mathcal{A}$. If $W \subset X$ is an open subset, it inherits an atlas consisting of pairs $(U \cap W, \phi|U \cap W)$ for each chart $(U, \phi) \in \mathcal{A}$. We call W an open submanifold of X when it is given this atlas. The domain of each chart is thus an open submanifold.

If Y is an m-dimensional manifold with atlas $\mathcal{B}$ and $f : X \to Y$ is a continuous map, f is said to be *smooth* if $\psi \circ f \circ \phi^{-1}$ is smooth, where defined, as a map from an open subset of $\mathbb{R}^n$ into $\mathbb{R}^m$ for each $(U,\phi) \in \mathcal{A}$, $(V,\psi) \in \mathcal{B}$. $f : X \to Y$ is a *diffeomorphism* if it is a homeomorphism and both f and f^{-1} are smooth.

In particular, taking $Y = \mathbb{R}$ or $\mathbb{C}(\cong \mathbb{R}^2)$, a function f on X is smooth if $f \circ \phi^{-1}$ is a smooth function on $\phi(U)$ for each chart (U,ϕ) on X. For example, the components x^μ of $x \in \mathbb{R}^{n+1}$ are smooth functions on $\mathbb{R}^{n+1}$, and their restrictions to S^n are smooth functions on S^n.

If X is a manifold, $C^\infty(X)$ will denote the space of smooth functions (real- or complex-valued according to context). It is an algebra since the result of addition, multiplication by scalars, or multiplication of smooth functions yields smooth functions.

If X and Y are manifolds with atlases $\mathcal{A}$, $\mathcal{B}$ of dimensions n and m, respectively, the Cartesian product $X \times Y$, which consists of pairs (x,y) with x in X, y in Y, can be given the structure of a manifold of dimension n + m by taking the obvious product atlas

$\mathcal{A} \times \mathcal{B} = \{(U \times V, \phi \times \psi) \mid (U,\phi) \in \mathcal{A}, (V,\psi) \in \mathcal{B}\}$. This we call the product manifold.

As an example, a *Lie group* is an abstract group G with the structure of a differentiable manifold such that the map defined by

$$(g_1, g_2) \mapsto g_1^{-1} g_2, \quad g_1, g_2 \in G,$$

from $G \times G$ to G is smooth. Then taking inverses, or translation on the left or right by elements of G, will yield diffeomorphisms of G to itself.

Let G be a Lie group and X a manifold, then a *smooth* (left) *action* of G on X is a smooth map from G x X to X, written

$$(g,x) \mapsto g \cdot x \quad , \quad g \in G, \quad x \in X,$$

such that

$$1 \cdot x = x, \quad g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x, \quad \text{for all } x \in X,\ g_1,\ g_2 \in G.$$

A right action is defined in the analogous fashion.

1.4. Let X be an n-dimensional manifold with atlas $\mathcal{Q}$. Then the components x^μ, $\mu = 1, \ldots, n$, of a point x in $\mathbb{R}^n$ can be considered as functions on $\mathbb{R}^n$, and as such are smooth. For each chart $a = (U,\phi)$ in $\mathcal{Q}$ we define *coordinate functions* $x^\mu_{(a)}$ on U by

$$x^\mu_{(a)} = x^\mu \circ \phi,$$

and $x^\mu_{(a)}$ is smooth on U. If f is a smooth function on X, $f \circ \phi^{-1}$ is a smooth function on $\phi(U)$. We can take its μth partial derivative $\partial_\mu(f \circ \phi^{-1})$ and transfer this back to U to obtain a function which we denote by $\partial^{(a)}_\mu f = (\partial_\mu(f \circ \phi^{-1})) \circ \phi$.

If $a = (U,\phi)$, $b = (V,\psi)$, are two charts on X, then we can apply this to $f = x^\mu_{(b)}$ on $U \cap V$ to obtain the Jacobian matrix $J_{(a,b)}$:

$$J_{(a,b)}{}^\mu{}_\lambda = \partial^{(b)}_\lambda \, x^\mu_{(a)} .$$

The chain rule for derivatives in $\mathbb{R}^n$ then implies

$$\partial^{(b)}_\lambda f = \sum_\mu \partial^{(a)}_\mu f \, J_{(a,b)}{}^\mu{}_\lambda \tag{1.1}$$

for any smooth function f on $U \cap V$. In particular, if $c = (W,\theta)$ is a third chart, we have

$$J_{(a,c)} = J_{(a,b)} \cdot J_{(b,c)}$$

on $U \cap V \cap W$, where the operation on the right hand side of this equation is the multiplication of matrices of functions.

1.5. The notation in the previous section has become cumbersome. For the rest of this Communication, where confusion will not arise, the label $a = (U,\phi)$ will be dropped from coordinates and derivatives, and explicit references to atlases, charts, and so on will be avoided. When we give a formula in terms of coordinates, it is to be understood that some chart has been chosen, and that the expression is relative to the coordinates of this chart.

1.6. Let $a = (U,\phi)$ be a chart on the n-dimensional manifold X and f a smooth function on X. We can arrange the functions $\partial^{(a)}_\mu f$ on U into a row vector which we denote by $df_{(a)}$. The collection $\{df_{(a)}\}_{a \in \mathcal{Q}}$ is denoted by df, and is called the *differential* of f. The transformation law (1.1) can be written in this notation as

$$df_{(b)} = df_{(a)} \, J_{(a,b)} .$$

The coordinates $x^\mu_{(a)}$ are themselves functions on U, so have differentials $(dx^\mu_{(a)})_{(a)}$, which we write as $dx^\mu_{(a)}$, and

$$(dx^\mu_{(a)})_\lambda = \delta^\mu_\lambda .$$

Then, for any smooth function f, we have the identity

$$df_{(a)} = \sum \partial^{(a)}_\mu f \, dx^\mu_{(a)} \quad , \quad \text{or} \qquad df = \sum \partial_\mu f \, dx^\mu$$

adopting the convention of the previous section.

We can generalize this as follows: A 1-form β on X is the assignment to each chart $a = (U,\phi)$ of a row vector $\beta_{(a)}$ of functions on U such that

$$\beta_{(b)} = \beta_{(a)} \, J_{(a,b)}$$

on $U \cap V$ for each pair of charts $a = (U,\phi)$, $b = (V,\psi)$. If the components of $\beta_{(a)}$ are $\beta^{(a)}_\mu$, $\mu = 1, \ldots, n$, then

$$\beta_{(a)} = \sum_\mu \beta^{(a)}_\mu \, dx^\mu_{(a)} .$$

Thus the differentials $dx^\mu_{(a)}$ form a basis for the 1-forms on U.

The functions on X we will also call 0-forms, and denote the vector space of functions by $\Omega^0(X)$. The space of 1-forms is denoted by $\Omega^1(X)$, and the operation of taking the differential gives a linear map

$$d : \Omega^0(X) \to \Omega^1(X)$$

which satisfies Leibnitz's Rule:

$$d(fg) = f\,dg + g\,df$$

for f, g in $\Omega^0(X)$.

There is an important generalization of this structure: the *exterior calculus*. For each integer $p \geq 0$ we have a space $\Omega^p(X)$ of p-*forms*. For $p \geq 2$ an element β of $\Omega^p(X)$ is an alternating covariant p-tensor. That is, to each chart $a = (U,\phi)$, we have an array $\beta^{(a)}_{\mu_1, \ldots, \mu_p}$ of smooth functions on U which is alternating under interchange of pairs of its suffices, and such that

$$\beta^{(b)}_{\lambda_1, \ldots, \lambda_p} = \sum_{\mu_1,\ldots,\mu_p} \beta^{(a)}_{\mu_1, \ldots, \mu_p} J_{(a,b)}{}^{\mu_1}{}_{\lambda_1} \cdots J_{(a,b)}{}^{\mu_p}{}_{\lambda_p} \tag{1.2}$$

on $U \cap V$ for any pair of charts $a = (U,\phi)$, $b = (V,\psi)$. p-forms may be added and multiplied by scalars or functions in the obvious, componentwise, fashion.

The *exterior derivative*

$$d : \Omega^p(X) \to \Omega^{(p+1)}(X)$$

is defined by

$$(d\beta)_{\mu_1, \ldots, \mu_{p+1}} = \sum_{i=1}^{p+1} (-1)^{i+1} \partial_{\mu_i} \beta_{\mu_1, \ldots, \widehat{\mu_i}, \ldots, \mu_{p+1}} \tag{1.3}$$

on each chart, where $\widehat{\mu_i}$ means that the suffix μ_i is omitted. Since β is alternating $d\beta$ is indeed a (p+1)-form, and one may verify that $d \circ d : \Omega^p(X) \to \Omega^{p+2}(X)$ is the zero map for all p.

Notice that $\Omega^p(X)$ consists of the zero element if $p > n$, since there can be no non-trivial alternating tensor whose degree exceeds the dimension.

A p-form β is said to be *closed* if $d\beta = 0$, and the closed p-forms form a subspace $Z^p(X)$ of $\Omega^p(X)$. If $p > 0$, a p-form β is said to be *exact* if there is a (p-1)-form γ with $\beta = d\gamma$. The exact p-forms form a subspace $B^p(X)$ of $\Omega^p(X)$, and, since $d(d\gamma) = d^2\gamma = 0$ for all (p-1)-forms γ, it follows $B^p(X) \subset Z^p(X)$.

The quotient vector space $H^p(X) = Z^p(X)/B^p(X)$ which consists of equivalence classes of closed p-forms, two p-forms being equivalent if their difference is exact, is the p-th de Rham *cohomology group* of X. It is also written $H^p(X; \mathbb{R})$ or $H^p(X; \mathbb{C})$ according as real- or complex-valued forms have been considered in its construction. It turns out that for a large class of manifolds (paracompact) the cohomology groups $H^p(X)$ depend only on the topology of X.

1.7. In this section we describe some further properties of differential forms and their derivatives.

A p-form β and a q-form γ may be multiplied to produce a (p+q)-form $\beta \wedge \gamma$. For $p, q \geq 1$ this is given by

$$(\beta \wedge \gamma)_{\mu_1, \ldots, \mu_{p+q}} = \frac{1}{p!q!} \sum_{\sigma \in P_{p+q}} \operatorname{sign}(\sigma)\, \beta_{\mu_{\sigma(1)}, \ldots, \mu_{\sigma(p)}} \gamma_{\mu_{\sigma(p+1)}, \ldots, \mu_{\sigma(p+q)}}$$

for each chart, where P_k denotes the permutation group on k letters, and $\operatorname{sign}(\sigma) = \pm 1$ is the signature of the permutation σ in P_k. Some authors use a different normalization factor ($\frac{1}{(p+q)!}$).

This multiplication is associative and graded commutative:

$$(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma) \quad ; \quad \beta \wedge \gamma = (-1)^{pq} \gamma \wedge \beta.$$

A straightforward calculation also shows

$$d(\beta \wedge \gamma) = (d\beta) \wedge \gamma + (-1)^p \beta \wedge (d\gamma).$$

If f is in $\Omega^0(X)$, β in $\Omega^p(X)$, we extend the above by defining

$$(f \wedge \beta)_{\mu_1, \ldots, \mu_p} = f \cdot \beta_{\mu_1, \ldots, \mu_p} = (\beta \wedge f)_{\mu_1, \ldots, \mu_p} .$$

Then for any p-form β we have, on each chart,

$$\beta = \sum_{\mu_1 < \ldots < \mu_p} \beta_{\mu_1, \ldots, \mu_p} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p} ,$$

$$= \frac{1}{p!} \sum_{\mu_1, \ldots, \mu_p} \beta_{\mu_1, \ldots, \mu_p} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p} ,$$

and

$$d\beta = \sum_{\mu_1 < \ldots < \mu_p} d\beta_{\mu_1, \ldots, \mu_p} \wedge dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p} .$$

The following lemma is fundamental, and its proof is very simple:

Lemma (Poincaré). Let $\beta \in Z^p(X)$, then each point x in X has a neighbourhood U with $\gamma \in \Omega^{p-1}(U)$ and $\beta = d\gamma$ on U.

Proof. Since any point x is in the domain of a chart, we can suppose $a = (U,\phi)$ is a chart with x in U and $\phi(U)$ an open unit ball centred at the origin in $\mathbb{R}^n$. Let $\beta_{\mu_1, \ldots, \mu_p} = \beta^{(a)}_{\mu_1, \ldots, \mu_p} \circ \phi^{-1}$ define functions on the unit ball, and set $\gamma^{(a)}_{\mu_1, \ldots, \mu_p} = \gamma_{\mu_1, \ldots, \mu_p} \circ \phi$ where

$$\gamma_{\mu_1, \ldots, \mu_{p-1}}(x^1, \ldots, x^n) = \sum_{\mu=1}^{n} x^\mu \int_0^1 t^{p-1} \beta_{\mu, \mu_1, \ldots, \mu_{p-1}}(tx^1, \ldots, tx^n)\, dt.$$

It is left to the reader to verify that γ defined by this equation has the desired property.

Q.E.D.

If $\beta \in Z^1(X)$ then $d\beta = 0$, whilst

$$d\beta_{\mu_1 \mu_2} = \partial_{\mu_1} \beta_{\mu_2} - \partial_{\mu_2} \beta_{\mu_1} .$$

Thus the Poincaré Lemma generalizes to p-forms the familiar result that

$$\partial_\mu \beta_\lambda = \partial_\lambda \beta_\mu$$

if and only if

$$\beta_\mu = \partial_\mu f$$

for some function f.

The Poincaré Lemma allows us to compute $H^p(\mathbb{R}^n)$. For $\mathbb{R}^n$ has a single neighbourhood $U = \mathbb{R}^n$ and $\phi(x) = x/(1 + |x|^2)^{\frac{1}{2}}$ whose image is the open unit ball. Thus $Z^p(\mathbb{R}^n) = B^p(\mathbb{R}^n)$ for all $p > 0$, and hence $H^p(\mathbb{R}^n) = 0$, $p > 0$. If f is in $H^o(\mathbb{R}^n) = Z^o(\mathbb{R}^n)$ then $df = 0$. Thus $\partial_\mu f = 0$ for all μ and hence f is a constant. Thus $H^o(\mathbb{R}^n; \mathbb{R}) = \mathbb{R}$. In particular, if X is any manifold (of dimension n) and U is an open set diffeomorphic to $\mathbb{R}^n$, and $d\beta = 0$, $\beta \in \Omega^p(X)$, $p > 0$, then $\beta = d\gamma$ on U for some $\gamma \in \Omega^{p-1}(U)$. In view of the previous remark that the cohomology groups depend only on the topology, it suffices to have U homeomorphic to $\mathbb{R}^n$ to obtain this result.

As a second example let us consider $H^1(S^n)$. We have $S^n = U_N \cup U_S$, and both U_N and U_S are diffeomorphic to $\mathbb{R}^n$ (see §1.2 for the notation). Let $\beta \in Z^1(S^n)$. Then there is a function $f_N(f_S)$ on $U_N(U_S)$ such that $\beta = df_N$ on U_N ($= df_S$ on U_S). Consider what happens on $U_N \cap U_S$. If $n > 1$, $U_N \cap U_S$ is connected, and we have

$$d(f_N - f_S) = df_N - df_S = \beta - \beta = 0.$$

Thus $f_N - f_S$ is constant, equal to c say, and then we define

$$f = \begin{cases} f_N & \text{on} \quad U_N \\ c+f_S & \text{on} \quad U_S \end{cases}$$

and obtain a smooth function on S^n with $df = \beta$. Thus $H^1(S^n) = 0$ for $n > 1$. For $n = 1$, $H^1(S^1) = \mathbb{R}$, which will follow from results to be described later.

A further useful operation on forms is the *pull-back*. If $f : X \to Y$ is a smooth map of manifolds and β is a p-form on Y, there is a p-form $f^*\beta$ on X called the pull-back of β by f. If we take a chart (V,ψ) on Y with coordinates y^μ, then $y^\mu \circ f$ is a smooth function on $f^{-1}(V)$, and on $f^{-1}(V)$ $f^*\beta$ is given by

$$f^*\beta = \sum_{\mu_1,\dots,\mu_p} \beta_{\mu_1,\dots,\mu_p} \circ f \; d(y^{\mu_1} \circ f) \wedge \dots \wedge d(y^{\mu_p} \circ f) \tag{1.4}$$

if

$$\beta = \sum_{\mu_1,\dots,\mu_p} \beta_{\mu_1,\dots,\mu_p} dy^{\mu_1} \wedge \dots \wedge dy^{\mu_p} .$$

The sets $f^{-1}(V)$ cover X as (V,ψ) runs through an atlas on Y, and the formulae (1.4) agree on overlaps, so giving $f^*\beta$.

A particular case is given by the inclusion map $i : U \hookrightarrow X$ of an open subset. The pull-back $i^* : \Omega^p(X) \to \Omega^p(U)$ is the *restriction* map and $i^*\beta$ is often written $\beta|U$. The pull-back is compatible with d and $\wedge$:

$$d(f^*\beta) = f^*(d\beta) \quad , \quad f^*(\beta \wedge \gamma) = (f^*\beta) \wedge (f^*\gamma).$$

1.8. A manifold X of dimension n is *orientable* if it has an atlas $\mathcal{A}^+$ such that all the Jacobian matrices $J_{a,b}$, a, b in $\mathcal{A}^+$ have positive determinant. An *orientation* of X is a maximal such atlas. It can be shown X is orientable if it has an n-form ω which vanishes nowhere (that is, for all charts $a = (U,\phi)$, x in U, and $\mu_1, \dots, \mu_n$, $\omega^{(a)}_{\mu_1,\dots,\mu_n}(x) \neq 0$). If ω is such a form we define $\mathcal{A}^+$ as all charts a with $\omega^{(a)}_{1,\dots,n}(x) > 0$ for all x in U. A given connected manifold X either has no orientation or two orientations. We shall consider only orientable manifolds, and for such manifolds we shall choose and fix an orientation. Then the manifold is said to be *oriented*.

If X is oriented and paracompact we can define the integral over X of n-forms ω, written

$$\int_X \omega$$

as follows: If ω has compact support contained in the domain of a chart $a = (U,\phi)$ in $\mathcal{A}^+$ then we set

$$\int_X \omega = \int_{\phi(U)} [\omega^{(a)}_{1,\dots,n} \circ \phi^{-1}] (x^1, \dots, x^n) \, dx^1 \dots dx^n .$$

For more general compactly supported forms we use a partition of unity with supports contained in coordinate patches. Details may be found in Kobayashi and Nomizu (1963).

If X is compact and oriented we may integrate any n-form. If β is an n-1 form then $d\beta$ is an n-form and

$$\int_X d\beta = 0.$$

This means that the map

$$\omega \mapsto \int_X \omega$$

from $\Omega^n(X)$ to $\mathbb{R}$ passes to the quotient to give a map

$$H^n(X;\ \mathbb{R}) \to \mathbb{R}\ .$$

This map is an isomorphism.

1.9. A *Riemannian metric* g on X assigns to each chart $a = (U,\phi)$ a positive-definite symmetric matrix $g^{(a)}_{\mu_1\mu_2}$ of real-valued functions on U such that, if $b = (V,\psi)$ is a second chart,

$$g^{(b)}_{\lambda_1\lambda_2} = \sum_{\mu_1,\mu_2} g^{(a)}_{\mu_1\mu_2}\, J_{(a,b)}{}^{\mu_1}{}_{\lambda_1}\, J_{(a,b)}{}^{\mu_2}{}_{\lambda_2}$$

on $U \cap V$. As usual $g^{\mu_1\mu_2}_{(a)}$ denotes the inverse matrix and $g^{(a)}$ the determinant of $g^{(a)}_{\mu_1\mu_2}$ on U.

g defines a bilinear form $\tilde{g}$ on $\Omega^p(X)$ for each $p \geq 1$ which takes its values in $\Omega^0(X)$. If β, γ are in $\Omega^p(X)$,

$$\tilde{g}(\beta,\gamma)|U = \frac{1}{p!} \sum_{\substack{\mu_1,\ldots,\mu_p \\ \lambda_1,\ldots,\lambda_p}} g^{\mu_1\lambda_1} \cdots g^{\mu_p\lambda_p}\, \beta_{\mu_1,\ldots,\mu_p}\, \gamma_{\lambda_1,\ldots,\lambda_p}\ .$$

If X is oriented on a chart U, there is also an n-form ρ on X, determined by g, called the *Riemannian volume*. If $a = (U,\phi)$ is a chart in $\mathcal{Q}$ then we set

$$\rho^{(a)}_{\mu_1,\ldots,\mu_n} = \sqrt{g^{(a)}}\ \mathrm{sign}\begin{pmatrix}1, & \ldots, & n \\ \mu_1, & \ldots, & \mu_n\end{pmatrix}.$$

It is easy to see that

$$\rho^{(b)}_{\lambda_1,\ldots,\lambda_n} = \rho^{(a)}_{\lambda_1,\ldots,\lambda_n}\ \mathrm{Det}\, J_{(a,b)}$$

on $U \cap V$ for two charts $a = (U,\phi)$, $b = (V,\psi)$, and this is the correct transformation rule for an n-form.

There is a unique map, the Hodge duality or star operator,

$$* : \Omega^p(X) \to \Omega^{n-p}(X)\ ,$$

such that

$$\tilde{g}(\beta,\gamma)\,\rho = \beta \wedge (*\gamma)$$

for any two p-forms β and γ. Explicitly

$$(*\beta)_{\mu_1,\ldots,\mu_{n-p}} = \frac{\sqrt{g^{(a)}}}{p!} \sum_{\substack{\mu_{n-p+1},\ldots,\mu_n \\ \lambda_1,\ldots,\lambda_p}} \mathrm{sign}\begin{pmatrix}\mu_1, & \ldots, & \mu_n \\ 1, & \ldots, & n\end{pmatrix} \beta_{\lambda_1,\ldots,\lambda_p}\, g^{\mu_{n-p+1}\lambda_1} \cdots g^{\mu_n\lambda_p}\ .$$

An explicit calculation shows that

$$*(*\beta) = (-1)^{p(n-p)}\,\beta$$

for β in $\Omega^p(X)$.

If λ is an everywhere strictly positive smooth function on X, then setting $g' = \lambda^2 g$ gives a new metric which is said to be *conformal* to g. Then, if $\beta,\gamma \in \Omega^p(X)$,

$$\tilde{g}'(\beta,\gamma) = \lambda^{-2p}\,\tilde{g}(\beta,\gamma)\ ,$$

and if ρ' is the Riemannian volume for g'

$$\rho' = \lambda^n \rho\ .$$

This means that the star operator for g' is $\lambda^{n-2p}*$. In particular, if the dimension n of X is even, say 2m, then $*$ acting on $\Omega^m(X)$ is conformally invariant. If, further, m is even, say n = 4k, then $**$ is the identity on $\Omega^{2k}(X)$. $*$ then has ± 1 as its eigenvalue. We let $\Omega^{2k}(X)^{\pm}$ be the corresponding eigenspaces. Elements of $\Omega^{2k}(X)^+$ are said to be *self-dual*, those of $\Omega^{2k}(X)^-$ *anti-self-dual*. If β is in $\Omega^{2k}(X)$ we can split it uniquely in the form

$$\beta = \beta^+ + \beta^-$$

with $\beta^{\pm}$ in $\Omega^{2k}(X)^{\pm}$. Then we have

$$\tilde{g}(\beta,\beta) = \tilde{g}(\beta^+,\beta^+) + \tilde{g}(\beta^-,\beta^-)$$

and

$$\tilde{g}(\beta,*\beta) = \tilde{g}(\beta^+,\beta^+) - \tilde{g}(\beta^-,\beta^-)\ .$$

If X is oriented and has a metric g, we can define an inner product on compactly supported forms by

$$(\beta,\gamma) = \int_X \tilde{g}(\beta,\gamma)\rho = \int_X \beta \wedge *\gamma\ .$$

There is a unique differential operator $\delta : \Omega^p(X) \to \Omega^{p-1}(X)$, the formal adjoint of d, satisfying

$$(d\beta,\gamma) = (\beta,\delta\gamma)\ , \qquad \beta \in \Omega^{p-1}(X), \qquad \gamma \in \Omega^p(X)\ .$$

It is easy to see that

$$\delta = (-1)^{np+n+1} *d* , \quad \delta^2 = 0.$$

Then

$$\Delta = d\delta + \delta d$$

acting on $\Omega^p(X)$ is called the *Laplacian* on p-forms, and a solution of

$$\Delta \beta = 0$$

is a *harmonic p-form*. Clearly $\Delta = (d + \delta)^2$ so is positive and self-adjoint. Further, $\Delta\beta = 0$ implies

$$0 = (\Delta\beta,\beta) = (d\beta, d\beta) + (\delta\beta, \delta\beta),$$

and hence

$$d\beta = 0 = \delta\beta.$$

If $\mathcal{H}^p(X)$ denotes the space of harmonic p-forms on X (it depends, of course, on the choice of metric) then $\mathcal{H}^p(X) \subset Z^p(X)$.

Theorem (Hodge). $\Omega^p(X) = d(\Omega^{p-1}(X)) + \mathcal{H}^p(X) + \delta(\Omega^{p+1}(X))$ is an orthogonal direct sum. $Z^p(X) = B^p(X) + \mathcal{H}^p(X)$ and hence $H^p(X) \cong \mathcal{H}^p(X)$ for all $p \geq 0$.

Since we already know $H^1(S^n) = 0$, we deduce there are no harmonic 1-forms on S^n, $n > 1$.

1.10. The notion of forms on a manifold X can be extended to include *vector-valued forms*. If V is a vector-space with a basis t_i, $i = 1, \ldots, N$, then a form with values in V consists of a linear combination $\beta = \sum_{i=1}^{N} \beta^i t_i$ where each β^i is itself just an ordinary form on X. If V has some extra structure, for example a metric or a Lie algebra bracket, this structure can be extended to the V-valued forms. A V-valued form $\beta = \sum_{i=1}^{N} \beta^i t_i$ is a p-form if β^i is in $\Omega^p(X)$ for all i. Let $\Omega^p(X) \otimes V$ denote the space of V-valued p-forms.

Suppose $V = \mathfrak{g}$ is a Lie algebra, then we can define an operation

$$\langle\cdot,\cdot\rangle : \Omega^p(X) \otimes \mathfrak{g} \times \Omega^q(X) \otimes \mathfrak{g} \to \Omega^{p+q}(X) \otimes \mathfrak{g}$$

extending simultaneously the exterior multiplication of forms and the Lie bracket of $\mathfrak{g}$. If $\beta = \sum_{i=1}^{N} \beta^i t_i$ is in $\Omega^p(X) \otimes \mathfrak{g}$ and $\gamma = \sum_{j=1}^{N} \gamma^j t_j$ is in $\Omega^q(X) \otimes \mathfrak{g}$ we define

$$\langle \beta, \gamma\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} \beta^i \wedge \gamma^j [t_i, t_j] .$$

This operation is neither a Lie bracket, nor an exterior multiplication, but is an example of a *graded Lie bracket*.

We can extend d to V-valued forms by setting

$$d(\sum \beta^i t_i) = \sum(d\beta^i)t_i .$$

Then, in the case $V = \mathfrak{g}$,

$$\left.\begin{aligned} &\langle\beta,\gamma\rangle = -(-1)^{pq} \langle\gamma,\beta\rangle , \quad d\langle\beta,\gamma\rangle = \langle d\beta,\gamma\rangle + (-1)^p \langle\beta,d\gamma\rangle , \\ &(-1)^{pr}\langle\beta, \langle\gamma,\delta\rangle\rangle + (-1)^{rq}\langle\delta, \langle\beta,\gamma\rangle\rangle + (-1)^{qp}\langle\gamma, \langle\delta,\beta\rangle\rangle = 0, \end{aligned}\right\} \quad (1.5)$$

where δ is in $\Omega^r(X) \otimes \mathfrak{g}$. The last equation above is the graded Jacobi identity.

An example of a $\mathfrak{g}$-valued form occurs if $\mathfrak{g}$ is the Lie algebra of the Lie group G; then $\omega = g^{-1}dg$ is a $\mathfrak{g}$-valued 1-form on G called the Maurer-Cartan form. It satisfies

$$d\omega + \tfrac{1}{2}\langle\omega,\omega\rangle = 0.$$

2. Vector Bundles

2.1. Let X be a smooth manifold. A real *vector bundle* of rank r over X is a smooth manifold E with a smooth map $\pi : E \to X$ such that each *fibre* $E_x = \pi^{-1}(x)$, $x \in X$, is a vector space of dimension r, and every point of X has a neighbourhood U with a diffeomorphism $\phi : \pi^{-1}U \to U \times \mathbb{R}^r$ such that

(i) $\pi_1 \circ \phi = \pi$,

(ii) $\phi : E_x \to \mathbb{R}^r$ is linear for each x in U.

Here $\pi_1 : U \times \mathbb{R}^r \to U$ denotes the projection on the first factor.

An elementary example is $E = X \times \mathbb{R}^r$ with $\pi = \pi_1$. This is the *product bundle*. Bundles can be *pulled back*: If $f : X \to Y$ is a smooth map and E a vector bundle over Y, we can define f^*E over X such that $(f^*E)_x = E_{f(x)}$. Details are left to the reader.

Two bundles (E,π), (F,τ) over X are *isomorphic* if there is a diffeomorphism $\phi : E \to F$ such that

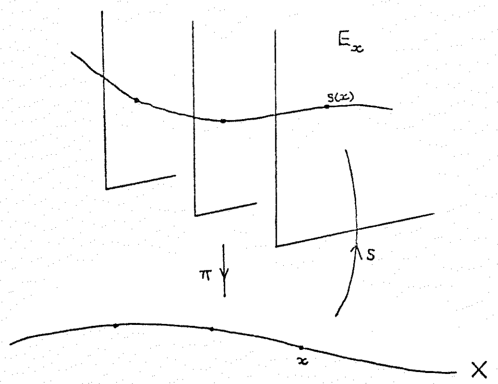
(i) $\tau \circ \phi = \pi$,

(ii) $\phi \mid E_x : E_x \to F_x$ is linear for all x in X.

A bundle is said to be *trivial* if it is isomorphic to the product bundle.

All the above makes sense over $\mathbb{C}$, and we obtain the notion of complex vector bundles.

A *section* s of a vector bundle (E,π) over X is a smooth map $s : X \to E$ with $\pi \circ s = \mathrm{id}_X$:



Sections may be added, multiplied by scalars and pointwise by smooth functions. We let ΓE denote the vector-space of sections of E.

If $U \subset X$ is an open subset, and (E,π) a vector bundle over X, the part $\pi^{-1}U$ lying over U gives a vector bundle $(E|U,\pi)$ called the *restriction* of E to U. Thus every point of X has a neighbourhood U with $E|U$ trivial. If $\phi : \pi^{-1}U \to U \times \mathbb{R}^r$ is an isomorphism with the product bundle, and $e_1, \ldots , e_r$ is the standard basis of $\mathbb{R}^r$, we obtain sections $s_1, \ldots , s_r$ of $E|U$ by

$$s_i(x) = \phi^{-1}(x,e_i) , \quad i = 1, \ldots , r.$$

Then $(s_1(x), \ldots , s_r(x))$ forms a basis for E_x for each x in U. We call this a *local frame* for E. Conversely, giving sections $s_i : U \to E$, $i = 1, \ldots , r$, with $(s_1(x), \ldots , s_r(x))$ a basis for E_x for each x in U, determines in a unique way an isomorphism of $E|U$ with $U \times \mathbb{R}^r$. If we have a second local frame $(t_1, \ldots , t_r)$ on an open set V, then we necessarily have

$$t_j(x) = \sum_{i=1}^{r} g_{ij}(x)\, s_i(x) , \quad x \in U \cap V$$

with

$$g : U \cap V \to GL(\mathbb{R}, r)$$

a smooth function.

Let (E,π) be a vector bundle over X, $\{U_\alpha\}$ an open covering of X such that on each U_α there is a local frame $(s_1^\alpha, \ldots , s_r^\alpha)$, then on $U_\alpha \cap U_\beta$ we have

$$s_j^\beta = \sum_{i=1}^{r} (g_{\alpha\beta})_{ij}\, s_i^\alpha \tag{2.1}$$

with

$$g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathbb{R}, r). \tag{2.2}$$

On $U_\alpha \cap U_\beta \cap U_\gamma$ we have

$$g_{\alpha\beta}\, g_{\beta\gamma}\, g_{\gamma\alpha} \equiv 1. \tag{2.3}$$

Also

$$g_{\alpha\beta}\, g_{\beta\alpha} \equiv 1, \quad g_{\alpha\alpha} \equiv 1 \tag{2.4}$$

on their common domains. Conversely, given a covering $\{U_\alpha\}$ of X and maps $g_{\alpha\beta}$ as in (2.2) satisfying (2.3) and (2.4), it is possible to construct a vector bundle (E,π) over X having local frames $(s_1^\alpha, \ldots, s_r^\alpha)$ on U_α satisfying (2.1) on $U_\alpha \cap U_\beta$. E is unique up to isomorphism. The maps $g_{\alpha\beta}$ are called the *transition functions* of E.

2.2. As a particular example, the covering of X by charts together with the Jacobian matrices of coordinate changes of §1.4 gives rise to a vector bundle T^*X, the *cotangent bundle* of X (there is also a *tangent bundle* TX obtained from ${}^tJ_{(a,b)}^{-1}$, but it does not concern us here). Then $\Omega^1(X) = \Gamma T^*X$. Further, $\Omega^p(X) = \Gamma\Lambda^pT^*X$ for a vector bundle Λ^pT^*X whose fibre at x is the p-th exterior power $\Lambda^p(T^*X_x)$.

The tensor products $E \otimes F$ of vector bundles (E,π), (F,σ) over X may be defined by means of the Kronecker product of their transition functions. There is a canonical identification of $(E \otimes F)_x$ with $E_x \otimes F_x$. We shall denote $\Gamma(\Lambda^pT^*X\otimes E)$ by $\Omega^p(X,E)$ (and ΓE by $\Omega^0(X,E)$ for consistency). An element β of $\Omega^p(X,E)$ may be identified with an assignment of an array $\beta^{(a)}_{\mu_1, \ldots, \mu_n}$ of sections of $E|U$ for each coordinate chart $a = (U,\phi)$ on X such that if $b = (V,\psi)$ is a second chart the transformation law (1.2) holds. The proof that these definitions are equivalent is left as an exercise for the reader. Using this identification there is an operation

$$\Omega^p(X) \times \Omega^q(X,E) \to \Omega^{p+q}(X,E)$$

which generalizes the exterior multiplication. Elements of $\Omega^p(X,E)$ are called *E-valued p-forms.*

We would like to define an operation on E-valued p-forms with properties similar to those of the ordinary exterior derivative. A *covariant derivative* D in E is a linear map

$$D : \Omega^0(X,E) \to \Omega^1(X,E)$$

such that

$$D(fs) = f\,Ds + df \wedge s,$$

for all f in $\Omega^0(X)$, s in $\Omega^0(X,E)$. There is a unique extension

$$D : \Omega^p(X,E) \to \Omega^{p+1}(X,E)$$

such that

$$D(\beta \wedge \gamma) = d\beta \wedge \gamma + (-1)^p \beta \wedge D\gamma$$

where β is in $\Omega^p(X)$ and γ in $\Omega^q(X,E)$. On a chart we can write

$$\beta = \sum_{\mu_1,\ldots,\mu_p} dx^{\mu_1}\wedge\ldots\wedge dx^{\mu_p} \wedge \beta_{\mu_1\wedge\ldots\wedge\mu_p}$$

with $\beta_{\mu_1, \ldots, \mu_p}$ in $\Omega^0(U,E)$ and then

$$D\beta = (-1)^p \sum_{\mu_1,\ldots,\mu_p} dx^{\mu_1}\wedge\ldots\wedge dx^{\mu_p} \wedge D(\beta_{\mu_1, \ldots, \mu_p}) .$$

We can compute D(Dfs) for any section s and smooth function f. Then

$$D(Dfs) = D(fDs + df \wedge s) = f\,D(Ds) + df \wedge Ds - df \wedge Ds + d^2f \wedge s$$
$$= f\,D(Ds).$$

Thus

$$D(Ds) = F_\wedge s$$

where F is in $\Omega^2(X, \mathrm{End}\,E)$. End E is the vector bundle over X whose fibre at x is the space $\mathrm{End}(E_x)$ of linear maps of E_x to itself. Hence $D^2 = 0$ only if $F = 0$. F is called the *curvature* of D, and if $F = 0$ the covariant derivative is said to be *flat*.

A metric in E is an inner product $(\cdot,\cdot)_x$ in each fibre E_x. (Hermitian if E is complex) such that, if s, t are smooth sections of E, then $(s(x), t(x))_x$ defines a smooth function $(s,t)(x)$ on X. If X is Riemannian and oriented then

$$\langle s, t\rangle = \int_X (s, t)\,\rho$$

makes the space Γ_cE of compactly supported smooth sections into a pre-Hilbert space. More generally, $\Lambda^pT^*X \otimes E$ has a metric derived from the Riemannian metric on T^*X together with the metric on E, and so by integration the compactly supported E-valued p-forms, $\Omega^p_c(X; E)$, become a pre-Hilbert space.

2.3. Let (E,π), (F,σ) be vector bundles over X and $P : \Gamma E \to \Gamma F$ a linear map. We say P is a *differential operator* (of order m) from E to F if

$$(Ps)(x) = \sum_{\substack{i,j\\|\alpha|\le m}} a(x)_{ij\alpha}\, \partial^\alpha f_j(x)\, t_i(x), \quad x \in U,$$

where

$$s(x) = \sum_j f_j(x)\, s_j(x) ,$$

$(s_1, \ldots, s_N)$ is a local frame for E, $(t_1, \ldots, t_M)$ a local frame for F, $\alpha = (\alpha_1, \ldots, \alpha_n)$, ($n = \dim X$), is an n-tuple of non-negative integers, $|\alpha| = \sum_{\mu=1}^n \alpha_\mu$, $\partial^\alpha = (-i\partial_1)^{\alpha_1} \ldots (-i\partial_n)^{\alpha_n}$, and U is the domain of some chart $a = (U,\phi)$.

If $\xi = (\xi_1, \ldots, \xi_n)$ is a vector in $\mathbb{R}^n$ we let

$$\sigma(P)(\xi,x)s(x) = \sum_{\substack{i,j \\ |\alpha|=m}} a(x)_{i,j,\alpha} \; \xi^\alpha f_j(x)\; t_i(x)$$

where $\xi^\alpha = \xi_1^{\alpha_1} \xi_2^{\alpha_2} \ldots \xi_n^{\alpha_n}$. If (ξ,x) is identified with a point of T^*X by $(\xi,x) \leftrightarrow \sum \xi_\mu dx^\mu$, then $\sigma(P)$ is invariantly defined as a section of $\mathrm{Hom}(\pi^*E,\pi^*F)$ where $\pi : T^*X \to X$ is the canonical projection. P is said to be *elliptic* if $\sigma(P)$ is an isomorphism of π^*E with π^*F (in particular E and F have the same rank).

Suppose now X is compact, Riemannian, and oriented, then if metrics are chosen in E and F we can take the pre-Hilbert space structures on ΓE and ΓF of §2.2 and complete them to obtain Hilbert spaces, denoted by $L^2(E)$, $L^2(F)$, respectively. If P is an elliptic differential operator from E to F, it defines an unbounded densely defined operator from $L^2(E)$ to $L^2(F)$, which is however Fredholm, along with the adjoint operator P^* from $L^2(F)$ to $L^2(E)$. P^* is the closure of a differential operator of order m (the formal adjoint of P) and both P and P^* have finite-dimensional kernels contained in ΓE and ΓF respectively. The *index* of P is defined to be

$$\mathrm{ind}\, P = \dim \ker P - \dim \ker P^*.$$

One of the most important results of 20th century mathematics is a formula for the index of P in terms of the symbol of P and the topological invariants of E, F and X. This is the celebrated *Atiyah-Singer Index Theorem:*

$$\mathrm{ind}\, P = (-1)^n\, (\mathrm{ch}[\sigma(P)]\,.\,\mathrm{td}[T^*X^{\mathbb{C}}])[T^*X]. \qquad (2.5)$$

The right-hand side of this equation is built from characteristic classes. (See the Appendix. We shall not go into details here, since in our application certain simplifications can be made on purely formal grounds. See Atiyah, Bott, Patodi (1973) for details on the Index Theorem.)

An apparent generalization can be made as follows. Suppose we have a sequence $E_0, E_1, E_2, \ldots, E_k$ of vector bundles over X and first order differential operators $d_i : \Gamma E_i \to \Gamma E_{i+1}$, $i = 0, \ldots, k-1$, with $d_{i+1} \circ d_i = 0$. Then

$$\Gamma E_0 \xrightarrow{d_0} \Gamma E_1 \xrightarrow{d_1} \Gamma E_2 \longrightarrow \ldots \xrightarrow{d_{n-1}} \Gamma E_k \qquad (2.6)$$

is called an *elliptic complex* if

$$\pi^*E_0 \xrightarrow{\sigma(d_0)} \pi^*E_1 \xrightarrow{\sigma(d_1)} \ldots \xrightarrow{\sigma(d_{k-1})} \pi^*E_k$$

is an exact sequence of vector bundles on $T^*X\backslash$(zero section). Equivalently, if we choose metrics in each E_i and consider

$$P : \Gamma \sum_{i\geq 0} E_{2i} \to \Gamma \sum_{i\geq 1} E_{2i-1}$$

given by $d_{2i} + d^*_{2i-1}$ on ΓE_{2i}, then P should be an elliptic operator. Applying the Index Theorem to P gives a formula for

$$\sum_{i\geq 0} (-1)^i \dim \mathrm{Ker}\, d_i \cap \mathrm{Ker}\, d^*_{i-1}\,.$$

Using Hodge theoretic results as in §1.9, $\mathrm{Ker}\, d_i \cap \mathrm{Ker}\, d^*_{i-1}$ is isomorphic to $\mathrm{Ker}\, d_i/\mathrm{Im}\, d_{i-1} = H^i$, which we call the i-th cohomology group of the complex (2.6). If $h^i = \dim H^i$ then the Index Theorem gives a formula for $\sum_{i\geq 0} (-1)^i h^i$.

In the case of interest we shall have $E_i = E \otimes L_i$, where E is a fixed vector bundle over X, the L_i are bundles built from T^*X, and $\sigma(d_i)$ is the identity on π^*E (acting from $\pi^*E \otimes \pi^*L_i$ to $\pi^*E \otimes \pi^*L_{i+1}$). Then the right-hand side of (2.1) is necessarily a linear function of $\mathrm{ch}\, E$. For $X = S^4$, only the components of $\mathrm{ch}\, E$ in degree 0 and 4 can be non-zero, and so for any complex of this kind on S^4 the Index formula necessarily has the form

$$a_0\, \mathrm{ch}_0\, E + a_2\, \mathrm{ch}_2\, E$$

and $\mathrm{ch}_0\, E = \mathrm{rk}\, E$. Here the numbers a_0, a_2 are independent of E. For a complex of length three ($k = 2$) we thus have

$$h^0 - h^1 + h^2 = a_0\, \mathrm{ch}_0\, E + a_2\, \mathrm{ch}_2\, E.$$

We will apply this formula in §4.2. For identification of the coefficients a_0, a_2 by direct calculation, see Atiyah et al. (1977,b).

3. Principal Bundles and Connections

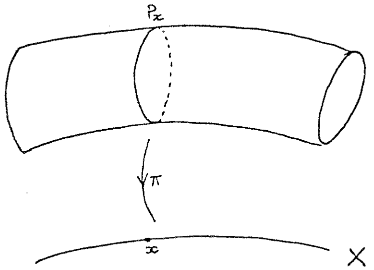
3.1. A fundamental and important object of study in differential geometry is the theory of principal bundles over a manifold having a given group G as structure group. In physics, these bundles describe how the internal spaces of a system with internal symmetry group G 'twist around' at different space-time points, in much the same way that coordinate transformations tell us how a curved space is pieced together. The principal bundle determines the topological structure of the gauge fields and Higgs fields of physics.

Mathematically, the definition goes as follows: Let X be a manifold and G a Lie group. A *principal G-bundle* over X, denoted (P,π,G), is another manifold P together with a smooth mapping π from P onto X with $\pi^* : \Omega^p(X) \to \Omega^p(P)$ injective (in mathematical language, π is a *submersion*), and a smooth action of G on P (which we write on the right: $p \cdot g$) with the properties

(i) $p \cdot g = p$ implies $g = 1$ (we say the action is *free*),

(ii) $\pi(p) = \pi(q)$ if and only if $q = p \cdot g$ for some g in G.

This definition may seem somewhat technical, but its main effect is that P may be viewed as a union of copies of G, one for each point of X, glued together in a smooth manner:

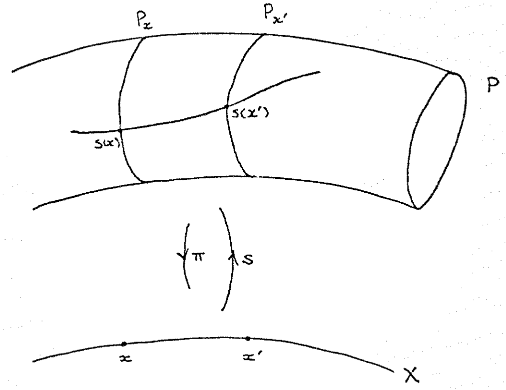


We think of P as lying above X, and π as the projection of P onto X. We denote by P_x the set of all points of P which are mapped by π onto a single point x in X. This is the 'fibre' of P over x. P_x and P_y are disjoint for $x \neq y$ and each is isomorphic with G as a set. We do not give P_x the structure of a group, and the relationship between P_x and G is the same as that between the orthonormal bases of a vector space at a point x, and the orthogonal group. There are precisely as many bases as elements of the group, but

no preferred way of identifying bases with group elements until we pick one basis and make comparisons relative to it.

One principal G-bundle always exists, namely the Cartesian product $P = X \times G$ with π the projection onto the first component. This is called the *trivial* or *product* bundle. One consequence of our requirement that π be a submersion is that locally every principal bundle looks like this. This justifies our claim that principal bundles are concerned with global properties, that is, with topology.

Let $U \subset X$ be open and (P,π,G) be a principal bundle over X. A *section* s over U is a smooth map $s : U \to P$ with $\pi(s(x)) = x$ for all x in U. Thus s assigns to each point in U an element of the fibre over that point. That is, it is a cross-section through the fibres over U, hence its name; see the picture.



In the part of P lying over U we now have a basepoint s(x) in each fibre P_x, so P_x is identified with G and $\pi^{-1}(U)$ with $U \times G$. A section thus sets up an isomorphism between a portion of P and the trivial bundle. This is the reason why we view sections only locally; the only principal bundle which admits a section defined on all of X is the trivial bundle.

Let $U \subset X$ be open and suppose we have two sections s, t of P over U. Then t(x) and s(x) lie in the same fibre P_x, so by condition (ii) there is g(x) in G with

$$t(x) = s(x) \cdot g(x),$$

and by (i) g(x) is uniquely determined by this equation. We thus obtain a function from

U into G. Such functions are known in physics as (non-abelian for general groups) *gauge transformations*. In this context they arise as transformations which change the reference point in a principal bundle. If we think of the principal bundle as representing bases in some internal degrees of freedom, defined over a curved space-time, then a section is a choise of basis at each point, and a gauge transformation arises as a change of this basis. Most principal bundles of interest may be viewed in this way.

On certain kinds of spaces all principal bundles are trivial. These are contractible spaces such as $\mathbb{R}^n$. Principal bundles over spaces of the form $X = \mathbb{R}^n \times Y$ are then determined by their restriction to Y, since they are constant in the $\mathbb{R}^n$ directions. As an example, consider $X = S^n$. As we have seen, this space can be covered by two open sets, U_N, U_S, each of which is diffeomorphic to $\mathbb{R}^n$. Thus any principal bundle (P,π,G) has a section s_N on U_N and another s_S on U_S. On the intersection $U_N \cap U_S$ we have two sections and hence a gauge transformation $g : U_N \cap U_S \to G$ with

$$s_N(x) = s_S(x) \cdot g(x), \quad x \in U_N \cap U_S .$$

This function g determines P completely. $U_N \cap U_S$ has the same topology as $\mathbb{R} \times S^{n-1}$, so that the topology of P is even determined by g as a function on S^{n-1}. Thus each principal bundle (P,π,G) on S^n has associated maps of S^{n-1} into G. It can be shown that P is essentially unchanged if this map is deformed continuously, so that in fact it is the homotopy group $\pi_{n-1}(G)$ (see Husemoller (1975) for full definitions) which determines the distinct principal bundles on S^n with structure group G.

This allows us to reduce the problem of studying principal bundles with a given structure group G on S^n for $n \geq 3$ to various special cases. For instance $\pi_{n-1}(G) = \pi_{n-1}(\tilde{G})$ where $\tilde{G}$ is a covering group of G, which means we may restrict attention to the case of G simply connected. Also $\pi_{n-1}(G) = \pi_{n-1}(K)$ where K is the maximal compact subgroup of G. Hence we can take G compact. Finally $\pi_{n-1}(G_1 \times G_2) = \pi_{n-1}(G_1) \times \pi_{n-1}(G_2)$ means we can treat each simple factor of the simply-connected, compact (and hence semi-simple) Lie group G separately. Thus, so far as the topology of principal bundles on S^n is concerned, we can assume G is a simply-connected, compact, simple Lie group.

The case in which we are interested is $n = 4$. It is known that if G is compact, simple, $\pi_3(G)$ is a single copy of the integers. It follows that each principal G-bundle

over S^4 determines an integer k, its *topological charge*, and this integer determines the bundle (at least up to a notion of isomorphism which need not concern us here). For a general group G, one integer will be obtained for each simple factor of its maximal compact subgroup.

SU(2) has the topology of S^3, so the gauge transformation g is essentially a map $g : S^3 \to S^3$. k is its winding number or degree. In Atiyah et al. (1977b) it is shown that every simple group G has an SU(2) subgroup such that g is determined by its restriction to SU(2) (at least the topology of (P,π,G) is determined), and k is the winding number of this map. Another way of defining k is given later.

3.2. We come now to the definition of the main objects of interest in Yang-Mills theory, the gauge potentials and gauge fields. The mathematical objects are known as connections and curvatures, respectively. We let (P,π,G) be a given principal bundle over a manifold X. A *connection* A in (P,π,G) is the assignment to each local section $s : U \to P$ of an element A_s of $\Omega^1(U) \otimes \mathfrak{g}$ (that is, a Lie algebra-valued 1-form) such that, if $t : U \to P$ is another section, related to s by $t = s \cdot g$ with $g : U \to G$ the corresponding gauge transformation, then

$$A_t = g^{-1} A_s g + g^{-1} dg . \tag{3.1}$$

In the physics literature, usually one fixed section is considered, and so the dependence of the gauge potential on the choice of the section is not mentioned. This can be done on Euclidean space $\mathbb{R}^n$ because there every bundle is trivial. On spaces such as S^n, which cannot be covered by a single coordinate chart, it is necessary to consider local gauge potentials and how they are related under changes of the local trivializations of the bundles in order to take into account the global topology of the situation.

The *curvature* or field strength F of the connection A in (P,π,G) is the assignment F_s of an element of $\Omega^2(U) \otimes \mathfrak{g}$ to each local section $s : U \to P$ determined by

$$F_s = dA_s + \tfrac{1}{2}\langle A_s, A_s \rangle \tag{3.2}$$

The full expression of F_s in terms of coordinates is

$$(F_s)^i_{\mu\nu} = \partial_\mu (A_s)^i_\nu - \partial_\nu (A_s)^i_\mu + \sum_{jk} c^i_{jk} (A_s)^j_\mu (A_s)^k_\nu \tag{3.3}$$

where c^i_{jk} are the structure constants of the Lie algebra $\mathfrak{g}$ determined by

$$[t_j, t_k] = \sum_i c^i_{jk} t_i .$$

It should be observed how economical is the expression (3.2) compared with (3.3) in terms of writing, and, more usefully, in terms of calculation. For it follows from $d^2 = 0$ and the identities (1.5) that

$$\begin{aligned} dF_s &= \tfrac{1}{2} d \langle A_s, A_s \rangle = \tfrac{1}{2} \langle dA_s, A_s \rangle - \tfrac{1}{2} \langle A_s, dA_s \rangle \\ &= \langle dA_s, A_s \rangle = \langle F_s, A_s \rangle . \end{aligned}$$

Thus

$$dF_s + \langle A_s, F_s \rangle = 0.$$

This is known as *Bianchi's Identity:* it is considerably less cumbersome to establish it thus than by using the expression (3.3).

If $t : U \to P$ is another section and $t = s \cdot g$ with $g : U \to G$ a gauge transformation, then it follows from (3.1) that

$$F_t = g^{-1} F_s g.$$

The field intensities are different kinds of objects from the potentials in that they do not have the inhomogeneous term $g^{-1}dg$ in their transformation rule. In fact F is in $\Omega^2(X,P(\mathfrak{g}))$ where $P(\mathfrak{g})$ is a vector bundle associated to P. This may be described as follows: If (P,π,G) is a principal bundle over X and $T : G \to \mathrm{End}\ V$ a representation of G, there is a vector bundle P(V) over X with fibre V obtained from $P \times V$ as the set of equivalence classes where (p_1,v_1) and (p_2,v_2) are equivalent if and only if there is g in G with

$$p_2 = p_1 \cdot g , \qquad v_2 = T(g^{-1})v_1 . \tag{3.4}$$

If we take $V = \mathfrak{g}$ with G acting by the adjoint representation, then we obtain $P(\mathfrak{g})$. Sections of $P(\mathfrak{g})$ may be described as follows: According to §3.1 a section ϕ chooses a point $\phi(x)$ in each fibre $P(\mathfrak{g})_x$ as x runs through X. Let $U \subset X$ be an open set with a section $s : U \to P$, then in the equivalence class determined by $\phi(x)$, $x \in U$, there is a pair of the form $(s(x), \phi_s(x))$ and so we obtain a function $\phi_s : U \to \mathfrak{g}$. According to the equivalence relation (3.4), if $t : U \to P$ is a second section with $t = s \cdot g$, $g : U \to G$, then we must have

$$\phi_t = g^{-1} \phi_s g \tag{3.5}$$

Thus a section of $P(\mathfrak{g})$ is, in physical terminology, a Higgs field in the adjoint representation.

We can also give a description of $\Omega^p(X,P(\mathfrak{g}))$ in terms of local sections $s : U \to P$. If (V,ψ) is a chart on U then $\beta \in \Omega^p(X,P(\mathfrak{g}))$ is given on V by the sections $\beta_{\mu_1, \ldots, \mu_p}$ of $P(\mathfrak{g})$, and by means of $s : U \to P$ these become functions $\beta_{\mu_1, \ldots, \mu_p, s} : V \to \mathfrak{g}$. Thus

$$\beta_s = \sum_{\mu_1,\ldots,\mu_p} \beta_{\mu_1, \ldots, \mu_p, s}\, dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p}$$

becomes an element of $\Omega^p(U) \otimes \mathfrak{g}$. Further, chasing through the identifications, we find

$$\beta_t = g^{-1} \beta_s g \tag{3.6}$$

if $t = s \cdot g : U \to P$ is a second local section. Conversely any family of $\mathfrak{g}$-valued p-forms satisfying (3.6) can be seen to give an element of $\Omega^p(X,P(\mathfrak{g}))$.

Now we can use the connection A in P to define a covariant derivative in $P(\mathfrak{g})$. This is the map $D : \Omega^0(X,P(\mathfrak{g})) \to \Omega^1(X,P(\mathfrak{g}))$ given, for any local section $s : U \to P$, by

$$(D\phi)_s = d \phi_s + \langle A_s, \phi_s \rangle .$$

This is easily seen to be a covariant derivative, and

$$D(D\phi) = \langle F, \phi \rangle ,$$

so that the curvature of D coincides with the curvature of A. Here, for $\beta \in \Omega^p(X,P(\mathfrak{g}))$, $\gamma \in \Omega^q(X,P(\mathfrak{g}))$, we define $\langle \beta,\gamma \rangle \in \Omega^{p+q}(X,P(\mathfrak{g}))$ by

$$\langle \beta,\gamma \rangle_s = \langle \beta_s, \gamma_s \rangle .$$

3.3. We continue with the situation of the previous section, and suppose in addition G is compact so that we can choose an invariant inner product on $\mathfrak{g}$ with respect to which the basis t_i is orthonormal. Let X have a Riemannian metric g, then we can define an inner product in $\Omega^p(X,P(\mathfrak{g}))$ with values in $\Omega^0(X)$ as follows: Consider $\sum_i \tilde{g}(\beta^i_s, \beta^i_s)$ for $s : U \to P$ a section. Since the inner product on $\mathfrak{g}$ is invariant we have

$$\sum_i \tilde{g} (\beta^i_s, \beta^i_s) = \sum_i \tilde{g}(\beta^i_t, \beta^i_t)$$

for any other section $t : U \to P$. This means that there is a function $|\beta|^2$ in $\Omega^0(X)$ such that, on an open U, we have

$$|\beta|^2 \mid U = \sum_i \tilde{g}(\beta_s^i, \beta_s^i)$$

for any section $s : U \to P$. If X is compact and orientable, and ρ denotes the Riemannian volume determined by g, we set

$$\|\beta\|^2 = \int_X |\beta|^2 \rho .$$

If F is the curvature of a connection A in (P,π,G) we can calculate $\|F\|^2$ in this way. This is the *total curvature*, or *Yang-Mills action*, of the field F. If F is a physical field F must minimize the action. Thus the problem of classical Yang-Mills theory is to study the minima, or more generally, the critical points of the functional $\|F\|^2$. $\|F\|^2$ can be formally defined when X is not compact, but then it need not be finite. One then needs to assume some growth behaviour on F in order that the minimization problem is well-defined.

The trivial bundle $P = X \times G$ has a connection A, known as the trivial connection, with the property that if $s_o(x) = (x, 1)$ is the constant section then $A_{s_o} = 0$. Then, if $s(x) = (x, g(x))$ is a general section, $A_s = g^{-1}dg$. Such a connection is said to be *pure gauge*. Note that for this connection $F_s = 0$ for all sections s. A connection A on a bundle (P,π,G) whose curvature F vanishes identically is said to be *flat*. If X is simply-connected this then implies P is isomorphic to the trivial bundle, and A is isomorphic to the trivial connection. Clearly, if A is flat, $\|F\| = 0$, so that this gives the absolute minimum.

However, the bundle (P,π,G) in which the connection A is defined is part of the information contained in the field, namely the topology of the field. This is characterized by discrete invariants (in the case $X = S^4$ and G simple a single integer will do), or topological charges, and cannot be varied continuously. Connections defined on bundles with different topological charges cannot be joined by a continuous curve of connections, so we should consider the restricted problem of minimizing $\|F\|^2$ amongst those connections A, defined on a single, fixed, principal bundle (P,π,G).

The Euler-Lagrange equations of the functional $\|F\|^2$, obtained by varying the potential A, are

$$D(*F) = 0. \qquad (3.7)$$

Here, we extend $*$ to bundle-valued forms, by

$$(*\beta)_s^i = *\beta_s^i .$$

In such generality solutions of this equation are difficult to find. If, however, X has dimension 4, $*F \in \Omega^2(X,P(\mathfrak{g}))$, so the possibility arises that $*F = \pm F$. In this case (3.7) becomes the Bianchi Identity, and is automatically fulfilled. In fact we can write $F = F^+ + F^-$ as in §1.9 and obtain

$$\|F\|^2 = \|F^+\|^2 + \|F^-\|^2 .$$

Suppose we are dealing with $X = S^4$, then (P,π,G) is determined by an integer k, which can be computed from F by the formula

$$8\pi^2 k = \|F^+\|^2 - \|F^-\|^2 ,$$

which is discussed in the Appendix. Since reversing the orientation changes the sign of $*$, and interchanges F^+ and F^-, we can restrict ourselves to a discussion of the case $k \geq 0$. Then

$$\|F\|^2 - 8\pi^2 k = 2\|F^-\|^2 \geq 0,$$

and so the theoretical absolute minimum of $\|F\|^2$ is $8\pi^2 k$ and occurs only if there is a connection A with $F^- = 0$, or equivalently $*F = F$. Such a connection is called *self-dual*. Thus for positive topological charge the absolute minimum of $\|F\|^2$ is given by a self-dual connection if it exists. The same argument shows that the anti-self-dual connections minimize $\|F\|^2$ for $k \leq 0$.

Recalling that every simple, simply-connected compact Lie group G contains SU(2) as a subgroup with the same topological charge, the existence of a self-dual connection with given charge for SU(2) establishes the existence of the absolute minimum for all non-abelian compact groups G. The first SU(2) self-dual connection was exhibited by Belavin et al. (1976), and had charge $k = 1$. Later 't Hooft found a 5k-parameter family of charge k for each $k \geq 1$, which was enlarged to a 13-parameter family for $k = 2$, and a $5k + 4$-parameter family for $k \geq 3$ by Jackiw et al.(1977). If G is abelian, $P(\mathfrak{g}) \cong X \times \mathfrak{g}$, and the curvature F of a connection in P is of the form $F = dA$, since $P = X \times G$ and $A \in \Omega^1(X) \otimes \mathfrak{g}$. Then $D{*F} = d(*dA)$, so equation (3.7) becomes

$$\delta\, dA = 0,$$

which implies

$$dA = 0,$$

and thus $F = 0$. Hence the only solution of equation (3.7) when G is abelian is the flat solution.

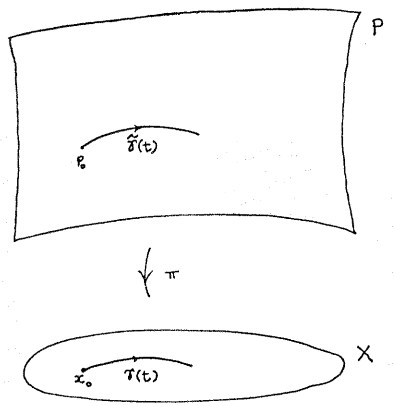
3.4. A connection A in a principal bundle (P,π,G) defines a notion of *parallel translation* between fibres of π. To describe this we need some more notation. A curve $\gamma(t)$ in X is a smooth map of an interval of $\mathbb{R}$ into X. In coordinates it is given by $x^\mu(t)$ with parameter t in $\mathbb{R}$. Let γ be a curve in X with $\gamma(0) = x_0$, and U be a neighbourhood of x_0 with a section $s : U \to P$. We let $A(t) = (A_s)_\mu(\gamma(t)) \dfrac{dx^\mu(t)}{dt}$ with respect to any coordinate neighbourhood containing $\gamma(t)$. The curve $A(t)$ in $\mathfrak{g}$ is independent of the coordinates chosen. Let $g(t)$ be a curve in G satisfying

$$\frac{dg(t)}{dt} + A(t)\, g(t) = 0, \qquad g(0) = 1. \tag{3.8}$$

It can be shown that this equation always has a unique solution with given initial value. Then we may define

$$\tilde{\gamma}(t) = s(\gamma(t)) \cdot g(t).$$

It follows that $\pi(\tilde{\gamma}(t)) = \gamma(t)$. If $p_0 = s(x_0)$, then $\tilde{\gamma}(0) = p_0$. We say $\tilde{\gamma}$ is the *horizontal lift* of $\gamma(t)$ starting at $p_0 \in \pi^{-1}(x_0)$.



If γ is not completely contained in a single open set U with a section $s : U \to P$, then we can cover γ by several open sets and piece together the horizontal lifts to give a unique curve $\tilde{\gamma}$ in P with $\tilde{\gamma}(0) = p_0 \in \pi^{-1}(x_0)$, $\pi\,\tilde{\gamma}(t) = \gamma(t)$, for all t. That the horizontal lift of γ does not depend on the open set U or section s is a consequence of the way A transforms, together with the uniqueness theorem for the differential equation (3.8).

If $p_0 \in \pi^{-1}(x_0)$ is given, we can always find a section s defined in some neighbourhood of x_0 with $s(x_0) = p_0$, so the horizontal lift $\tilde{\gamma}$ of a curve with $\gamma(0) = x_0$ can always be made to have $\tilde{\gamma}(0) = p_0$ for the given point p_0. If $\gamma(1) = \gamma(0) = x_0$, γ is a closed loop in X, but the horizontal lift $\tilde{\gamma}$ need not be closed. It must however return to a point of the fibre over x_0, so there is an element g_γ of G with

$$\tilde{\gamma}(1) = \tilde{\gamma}(0) \cdot g_\gamma = p_0 \cdot g_\gamma .$$

Let H_{p_0} be the subset of G consisting of all the group elements g_γ as γ varies over all closed curves beginning and ending at x_0. Then H_{p_0} is a subgroup of G (in fact a Lie subgroup), and is called the *holonomy group* of the connection at p_0.

It is not hard to show that

$$H_{p_0 \cdot g} = g^{-1} H_{p_0}\, g.$$

It is left as an exercise for the reader to show that (assuming X is connected) the holonomy groups H_{p_0}, H_{p_1} are conjugate for any pair of points p_0, p_1 in P. H_{p_0} is not necessarily connected. Its identity component is obtained by considering curves γ which may be shrunk smoothly to a point. If X is simply-connected this is always possible and hence H_{p_0} is connected in this case.

3.5. Let (P,π,G) be a principal bundle over X and H a closed subgroup of G. The principal bundle is said to be *reducible* to H if there is a submanifold $Q \subset P$ such that $(Q,\pi|Q,H)$ is a principal bundle over X. If A is a connection on P and $(Q,\pi|Q,H)$ is a reduction of (P,π,G), we say the connection A reduces to $(Q,\pi|Q,H)$ if, for local sections $s : U \to Q$, A_s is an $\mathfrak{h}$-valued 1-form, where $\mathfrak{h} \subset \mathfrak{g}$ is the Lie algebra of H.

The *holonomy reduction theorem* says that a connection A and the principal bundle (P,π,G) on which it is defined can always be reduced to the holonomy group H_{p_0}. In this case Q consists of all points p in P which can be joined to p_0 by some horizontal lift $\tilde{\gamma}$

of a curve γ with $\gamma(0) = \pi(p_o)$, $\gamma(1) = \pi(p)$.

If $G = H_{p_o}$ the connection is said to be *irreducible*. Let $\mathfrak{h}_{p_o}$ be the Lie algebra of the holonomy group H_{p_o}. Then $\mathfrak{h}_{p_o}$ is given by the Ambrose-Singer theorem in terms of the curvature. The precise statement may be found in Ambrose & Singer (1953), and a statement in local coordinates in Loos (1967). For the general theory of connections see Kobayashi & Nomizu (1963).

3.6. If the curvature F vanishes, it follows from the Ambrose-Singer theorem that $\mathfrak{h}_{p_o} = 0$, and if X is simply-connected $H_{p_o} = \{1\}$. Then Q consists of a single element from each fibre, and so is the image of a global section $s : X \to P$; and since A_s takes its values in $\mathfrak{h}_{p_o}$ we have $A_s = 0$. If $t : U \to P$ is any local section we have $t = s \cdot g$ on U, so

$$A_t = g^{-1}\, dg. \tag{3.9}$$

Thus the vanishing of the curvature of the connection implies that the connection forms A_t are a pure gauge (see §3.3).

3.7. Let (P,π,G) be a principal bundle over X with connection A and $T : G \to \mathrm{End}\, V$ a representation of G; then there is a covariant derivative D acting on sections of P(V) to produce elements of $\Omega^1(P(V))$. If ψ is a section of P(V) it is said to be *covariant constant* if $D\psi = 0$. In coordinates this means

$$\partial_\mu \psi_s + T((A_s)_\mu)\, \psi_s = 0$$

for a section $s : U \to P$ (ψ_s is defined in the same way as for the adjoint representation). If $\gamma(t)$ is a curve, given in coordinates by $x^\mu(t)$, we have

$$\frac{d}{dt} \psi_s(\gamma(t)) + T((A_s)_\mu \frac{dx^\mu(t)}{dt})\, \psi_s = 0\ ,$$

and hence

$$\frac{d}{dt} T(g(t)^{-1})\, \psi_s(\gamma(t)) = 0,$$

where g(t) satisfies (3.8). Integrating, we obtain

$$\psi_s(\gamma(t)) = T(g(t))\, \psi_s(\gamma(0)). \tag{3.10}$$

This means that a covariant constant section ψ is determined by its value at one point x_o, say, and for every closed curve γ at x_o we have

$$\psi_s(x_o) = T(g_\gamma)\, \psi_s(x_o).$$

Conversely, if we have a vector v in V with

$$v = T(g)\, v \tag{3.11}$$

for all g in H_{p_o}, there is a unique covariant constant section ψ, with $\psi(x_o) = v$ determined by (3.10). Let $V^{H_{p_o}}$ denote all v in V satisfying (3.11) for all g in H_{p_o}. In the case where $V = \mathfrak{g}$ and T is the adjoint representation, $\mathfrak{g}^{H_{p_o}}$ consists of all elements of the Lie algebra of G which commute with elements of H_{p_o} (or $\mathfrak{h}_{p_o}$ if H_{p_o} is connected).

4. THE MODULI OF SELF-DUAL CONNECTIONS

4.1. Let (P,π,G) be a principal bundle over X. An automorphism of the bundle is a diffeomorphism $\tau : P \to P$ such that

(i) $\pi \circ \tau = \pi$;

(ii) $\tau(p \cdot g) = \tau(p) \cdot g$, for all $p \in P$, $g \in G$.

(i) means that τ maps each fibre to itself, and (ii) that it commutes with the action of G on each fibre. We let $\mathcal{G}_P$ be the group of all automorphisms of the bundle P. Let A be a connection on P and $s : U \to P$ be a local section. Then $\tau \circ s$ is also a local section and we may define

$$(\tau^*A)_s = A_{\tau \circ s}.$$

We leave it to the reader to check that τ^*A is again a connection in P.

If τ is in $\mathcal{G}_P$ and p in P then $\tau(p)$ is in the same fibre as p by (i), so there is $g_\tau(p)$ in G with

$$\tau(p) = p \cdot g_\tau(p). \tag{4.1}$$

This defines a smooth map

$$g_\tau : P \to G,$$

and from (ii) it follows that

$$g_\tau(p \cdot g) = g^{-1} g_\tau(p)\, g, \quad \text{for all } p \in P,\ g \in G. \tag{4.2}$$

Conversely, given a smooth map $g_\tau : P \to G$ satisfying (4.2), then (4.1) defines an automorphism τ of P. Note that $(\tau \circ s)(x) = \tau(s(x)) = s(x) \cdot g_\tau(s(x))$, and thus

$$(\tau^*A)_s = (g_\tau \circ s)^{-1} A_s\, g_\tau \circ s + (g_\tau \circ s)^{-1}\, d(g_\tau \circ s).$$

Hence τ^*A is obtained from A by gauge transformations, and for physical purposes τ^*A and A may be identified.

We let $\mathcal{A}_P$ denote the space of all connections in P, and we let $\mathcal{G}_P$ act on $\mathcal{A}_P$ by

$$\tau \cdot A = (\tau^{-1})^*A;$$

then the space of orbits $\mathcal{G}_P \backslash \mathcal{A}_P = \mathcal{C}_P$ represents the set of gauge inequivalent connections on P.

Now suppose X is four-dimensional, Riemannian, and oriented, then those connections with self-dual curvature form a subset $\mathcal{S}_P$ of $\mathcal{A}_P$ which is invariant under $\mathcal{G}_P$, and we

denote the set of orbits in $\mathcal{S}_P$ by $\mathcal{M}_P$. $\mathcal{M}_P$ is the *moduli space* of self-dual connections on P. If X is S^4 and G simple (which we assume from now on), then P is determined by an integer $k \geq 0$, and we denote $\mathcal{A}_P$ by $\mathcal{A}_k$ and so on.

It is not hard to see that A and τ^*A have conjugate holonomy groups, so that it makes sense to speak of the irreducible elements of $\mathcal{M}_k$ which form a subset $\mathcal{M}_k^{o}$.

Theorem (Atiyah et al., 1977b). $\mathcal{M}_k^{o}$ is a manifold of dimension $a_G k - \dim G$ where a_G is an integer determined by the root structure of $\mathfrak{g}$, for k sufficiently large.

For $G = SU(2)$, $a_G = 8$, and for a complete list and more details, see Atiyah et al. (1977b). The remainder of this chapter is devoted to a sketch of the proof of this theorem.

4.2. Let A, A' be two connections in (P,π,G). For each local section $s : U \to P$, the difference

$$a_s = A'_s - A_s$$

is an element of $\Omega^1(U) \otimes \mathfrak{g}$. If $t = s \cdot g$ is a second section,

$$a_t = A'_t - A_t = g^{-1}A'_s g + g^{-1}dg - (g^{-1}A_s g + g^{-1}dg)$$

$$= g^{-1}a_s g.$$

Thus the difference of two connections defines an element of $\Omega^1(X, P(\mathfrak{g}))$. Put differently, fixing one connection A in P, every other connection has the form $A + a$ with a in $\Omega^1(X,P(\mathfrak{g}))$. Hence sets $\mathcal{A}_P$ and $\Omega^1(X,P(\mathfrak{g}))$ are bijective. A straightforward calculation shows that if F is the curvature of A then $A + a$ has curvature

$$F + Da + \tfrac{1}{2}\langle a, a\rangle,$$

where D is the covariant exterior derivative determined by A. If A is self-dual, then $A + a$ is also self-dual, if and only if

$$(Da)_- + \tfrac{1}{2}\langle a, a\rangle_- = 0. \tag{4.3}$$

Let $D_-a = (Da)_-$, then D_- is the differential operator from $T^*X \otimes P(\mathfrak{g})$ to $\Lambda^2_- T^*X \otimes P(\mathfrak{g})$. (4.3) is a nonlinear differential equation whose solutions are the points of $\mathcal{S}_P$, and we want to find the solutions of (4.3) up to bundle automorphisms.

The idea is to linearize the problem around a given solution, solve the linearized problem, and then use techniques of deformation theory to deduce the corresponding solution to the nonlinear problem.

Let $A + a(t)$ be a curve of connections through A. Then $a(0) = 0$, and the tangent to this curve is given by $a'(0) \in \Omega^1(X,P(\mathfrak{g}))$. The curvature of $A + a(t)$ is

$$F + Da(t) + \tfrac{1}{2}\langle a(t), a(t)\rangle$$

and this is self-dual if

$$D_-a(t) + \tfrac{1}{2}\langle a(t), a(t)\rangle_- = 0.$$

Differentiating and putting t equal to 0, we obtain

$$D_-a'(0) = 0$$

as the linearized condition for self-duality. Thus the space of tangents to curves of self-dual connections coincides with the kernel of the linear differential operator

$$D_- : \Omega^1(X,P(\mathfrak{g})) \to \Omega^2_-(X,P(\mathfrak{g})).$$

We want to know which connections are obtained from A by bundle automorphisms. Suppose τ_t is a curve of bundle automorphisms, with τ_0 the identity. This corresponds with a curve of maps g_{τ_t} of P to G satisfying

$$g_{\tau_t}(p \cdot g) = g^{-1}\, g_{\tau_t}(p)\, g \qquad \text{for all } p \in P,\ g \in G.$$

The tangent to this curve at 0 will be a map $\phi : P \to \mathfrak{g}$ satisfying

$$\phi(p \cdot g) = g^{-1}\, \phi(p)\, g \qquad \text{for all } p \in P,\ g \in G.$$

If we put

$$\phi_s = \phi \circ s$$

we see that ϕ defines a Higgs field: $\phi \in \Omega^0(X,P(\mathfrak{g}))$. Further, if

$$A + a(t) = \tau_t{}^*A,$$

then

$$a(t)_s = (g_{\tau_t} \circ s)^{-1} A_s (g_{\tau_t} \circ s) + (g_{\tau_t} \circ s)^{-1} d\, g_{\tau_t} \circ s$$

for local sections $s : U \to P$. Differentiating with respect to t, and putting $t = 0$, we find

$$a'(0)_s = -[\phi_s, A_s] + d\phi_s\ ,$$

or

$$a'(0) = D\phi\ .$$

Thus the tangents to curves which are equivalent to A by bundle automorphisms are the images of

$$D : \Omega^0(X,P(\mathfrak{g})) \to \Omega^1(X,P(\mathfrak{g})).$$

It follows that, formally, the space of tangents to $\mathcal{M}_k$ at the orbit of A is the cohomology group H^1 of the complex

$$\Omega^0(X,P(\mathfrak{g})) \xrightarrow{D} \Omega^1(X,P(\mathfrak{g})) \xrightarrow{D_-} \Omega^2_-(X,P(\mathfrak{g})).$$

This complex is elliptic, so the Index Theorem of Chapter 2 gives a formula for $h^0 - h^1 + h^2$. If we can calculate h^0 and h^2, we will know how large h^1 is.

Now the vector bundles E_i have the form $E \otimes L_i$ with L_0 trivial, $L_1 = T^*X$, $L_2 = \Lambda^2_-T^*X$, and $E = P(\mathfrak{g})$. Thus the topological formula for the index is linear in $chP(\mathfrak{g})$ and hence

$$h^0 - h^1 + h^2 = ak + b \dim \mathfrak{g}\ ,$$

because (as we see from the Appendix) $ch_2P(\mathfrak{g})$ is proportional to k, and $ch_0P(\mathfrak{g})$ is always the rank of the bundle, in this case $\dim \mathfrak{g}$.

If we take $k = 0$, we have $F = 0$; the bundle $P(\mathfrak{g})$ is trivial, isomorphic to $X \times \mathfrak{g}$ and so $H^0 = H^0(S^4) \otimes \mathfrak{g}$, $H^1 = H^1(S^4) \otimes \mathfrak{g}$, $H^2 \subset H^2(S^4) \otimes \mathfrak{g}$. But $H^1(S^4) = H^2(S^4) = 0$, and $H^0(S^4) = \mathbb{R}$. Thus $h^0 = \dim \mathfrak{g}$, $h^1 = h^2 = 0$ in this case. Thus $b = 1$.

Consider H^0 for $k > 0$. This is the kernel of $D : \Omega^0(X,P(\mathfrak{g})) \to \Omega^1(X,P(\mathfrak{g}))$ which was shown in Chapter 3 to be isomorphic to the dimension of the centralizer of the holonomy group. If A is irreducible then $H^0 = 0$. In Atiyah et al.(1977b), Bernard et al. (1977), and Schwarz (1977) it is shown that H^2 can be identified with the kernel of an operator of the form $\Delta + {}^R/_3$, where Δ is the Laplace operator on a suitable bundle, and R is the scalar curvature of S^4. Since $R = 12 > 0$, $\Delta + {}^R/_3$ has no kernel, so $h^2 = 0$ for all groups G and all $k > 0$. Hence, if A is irreducible, we have

$$h^1 = -ak - \dim \mathfrak{g}\ .$$

For $G = SU(2)$ and $k = 1$, it is shown in Atiyah et al. (1977b) that the space $\mathcal{M}_1$ $(= \mathcal{M}_1^0)$ is hyperbolic 5-space, so

$$5 = -a1 - 3,$$

or

$$a = -8.$$

Thus for $G = SU(2)$

$$h^1 = 8k - 3.$$

There is a similar result for each simple group, given in the references cited above.

4.3. In order to show that $\mathcal{M}_k$ (for G = SU(2)) actually has dimension 8k - 3, it is necessary to show that every element of H^1 is actually the tangent to some curve of solutions of the non-linear equation (4.3) and that (at least locally) 8k - 3 parameter families exist with no equivalences amongst them, by means of bundle automorphisms. In Atiyah et al. (1977b) techniques of deformation theory, first developed by Kuranishi (1965), are used to establish this, and to show that $\mathcal{M}_k^o$ is a Hausdorff manifold for all k and all G. For a given G, these authors also determine which values of k have $\mathcal{M}_k^o \neq \emptyset$.

5. The Construction of self-dual Connections

5.1. We know from the main result of Chapter 4 that the dimension of the moduli space of self-dual SU(2) connections on S^4 of charge k is 8k - 3. However, the largest family of solutions constructed by means of an ansatz is that of Jackiw et al. (1977) of dimension 5k + 4. The first step in finding a construction for all self-dual SU(2) connections was taken by Atiyah & Ward (1977), who used the twistor transform of Penrose (1977) to translate the problem into one of algebraic geometry.

Atiyah, Hitchin, Manin & Drinfeld (1978) were able to apply the results of Horrocks (1964) and Barth & Hulek (1978) to give a construction of all solutions. It is not possible in this Communication to cover all the material necessary for a full treatment of their results. In this final Chapter we shall sketch some of the main steps involved. The theory of complex manifolds, sheaves, and algebraic geometry needed for a proper understanding of the proofs may be found in Morrow & Kodaira (1971), Wells (1977), Chern (1967), Griffiths & Adams (1974) and Hartshorne (1977).

5.2. If, instead of taking $\mathbb{R}^n$ as the image space for coordinate maps as we did in Chapter 1, we take $\mathbb{C}^n$, and we require the transition maps between different charts to be *holomorphic* (that is, expandable in convergent power series about every point where they are defined), then we obtain the notion of an n-dimensional *complex manifold*. The most elementary example is $\mathbb{C}^n$ itself.

A family of complex manifolds of especial interest for us is the family of *projective spaces* $P^n(\mathbb{C})$, n = 1, 2, 3, ... $P^n(\mathbb{C})$ is the space of complex one-dimensional subspaces (or *lines*) in $\mathbb{C}^{n+1}$. If z is a non-zero vector in $\mathbb{C}^{n+1}$ we let [z] denote the line which it spans:

$$[z] = \{cz \in \mathbb{C}^{n+1} \mid c \in \mathbb{C}\} .$$

Let $z = (z_o, \dots , z_n)$ and

$$U_i = \{ [z] \in P^n(\mathbb{C}) \mid z_i \neq 0\} .$$

Then $U_o, \dots , U_n$ are open subsets of $P^n(\mathbb{C})$ which form a covering. We define

$$\phi_i : U_i \to \mathbb{C}^n$$

by

$$\phi_i([z]) = (z_0/z_i, \dots, z_{i-1}/z_i, z_{i+1}/z_i, \dots, z_n/z_i)$$

and obtain a family $\{(U_i, \phi_i)\}_{i=0}^n$ of charts on $P^n(\mathbb{C})$ which, as the reader may check, forms an atlas for $P^n(\mathbb{C})$ as a complex manifold.

$P^1(\mathbb{C})$ is known as the complex projective line, or Riemann sphere. It has the topology of S^2. $P^1(\mathbb{C})$ is the space of lines in $\mathbb{C}^2$. If $V \subset \mathbb{C}^{n+1}$ is any two-dimensional subspace, those lines in $\mathbb{C}^{n+1}$ which actually lie in V form a subset L of $P^n(\mathbb{C})$ which is clearly isomorphic to $P^1(\mathbb{C})$. L is said to be a (projective) line in $P^n(\mathbb{C})$. $P^3(\mathbb{C})$ contains a special family of lines parametrized by S^4 which is the basis for the results of Atiyah & Ward (1977). This is most easily described in terms of the quaternions $\mathbb{H}$.

The quaternions are a non-commutative algebra spanned by an identity 1, and three anti-commuting elements i, j, k with

$$i^2 = j^2 = k^2 = -1, \quad ij = k.$$

A quaternion q is a linear combination

$$q = q^1 1 + q^2 i + q^3 j + q^4 k,$$

and may be written

$$q = (q^1 1 + q^2 i) + j(q^3 1 - q^4 i).$$

If we identify the quaternions of the form

$$q^1 1 + q^2 i$$

with the corresponding elements of the complex numbers $\mathbb{C}$, then any quaternion q determines two complex numbers z_1, z_2 by

$$q = z_1 + jz_2$$

and a pair of quaternions (q_1, q_2) in $\mathbb{H}^2$ determines a point (z_1, z_2, z_3, z_4) in $\mathbb{C}^4$ by

$$q_1 = z_1 + jz_2, \qquad q_2 = z_3 + jz_4 . \tag{5.1}$$

A line $[z]$ in $\mathbb{C}^4$ determines a subset

$$\{(q_1 c, q_2 c) \in \mathbb{H}^2 \mid c \in \mathbb{C}\}$$

in $\mathbb{H}^2$ which can clearly be identified with $[z]$, where q_1, q_2 are determined by (5.1). The subset

$$\{(q_1 c, q_2 c) \in \mathbb{H}^2 \mid c \in \mathbb{H}\}$$

is a quaternionic line in $\mathbb{H}^2$, and the set of all these lines is denoted by $P^1(\mathbb{H})$. We thus have a map

$$\pi : P^3(\mathbb{C}) \to P^1(\mathbb{H}) \tag{5.2}$$

which sends any complex line in $\mathbb{C}^4$ to the corresponding quaternionic line in $\mathbb{H}^2$.

Let c in $\mathbb{H}$ be written

$$c = c_1 + jc_2$$

with c_1, c_2 in $\mathbb{C}$, and consider

$$(q_1 c, q_2 c) = (z_1 c_1 - \bar{z}_2 c_2 + j(z_2 c_1 + \bar{z}_1 c_2), z_3 c_1 - \bar{z}_4 c_2 + j(z_4 c_1 + \bar{z}_3 c_2)).$$

As c varies we obtain all linear combinations of

$$(z_1, z_2, z_3, z_4), \quad (-\bar{z}_2, \bar{z}_1, -\bar{z}_4, \bar{z}_3).$$

These two vectors are always linearly independent and hence define a plane in $\mathbb{C}^4$, or equivalently a line in $P^3(\mathbb{C})$. Thus the fibres of the map π (see (5.2)) are lines in $P^3(\mathbb{C})$. If we can demonstrate that $P^1(\mathbb{H}) = S^4$, then we have the desired fibering.

S^4 is the unit sphere in $\mathbb{R}^5$. We may identify $\mathbb{R}^5$ with $\mathbb{H} \times \mathbb{R}$ so that

$$S^4 = \{(q,r) \in \mathbb{H} \times \mathbb{R} \mid |q|^2 + r^2 = 1\} .$$

Then if $[q_1, q_2]$ denotes a quaternionic line in $\mathbb{H}^2$, we map it to

$$\left(\frac{2q_1\bar{q}_2}{|q_1|^2 + |q_2|^2} , \frac{|q_1|^2 - |q_2|^2}{|q_1|^2 + |q_2|^2} \right) \in \mathbb{H} \times \mathbb{R} .$$

This is clearly in S^4, and gives the equivalence of $P^1(\mathbb{H})$ with S^4.

The open set U_N (see §1.2) corresponds with the set of pairs $[q_1, q_2]$ with $q_2 \neq 0$. This may be mapped to $\mathbb{H}$ by

$$[q_1, q_2] \mapsto 2q_1 q_2^{-1} ,$$

and this coincides with the coordinate map ϕ_N when we identify $\mathbb{H}$ with $\mathbb{R}^4$. Similarly U_S corresponds with $q_1 \neq 0$ and ϕ_S with

$$[q_1, q_2] \mapsto 2q_2 q_1^{-1} .$$

If β is a form on S^4 we can pull it back (§1.7) to $P^3(\mathbb{C})$ by means of the map π, regarding $P^3(\mathbb{C})$ as a C^∞-manifold of six dimensions. We shall see what effect this has on 2-forms. Let β be a 2-form, and $U_4 \subset P^3(\mathbb{C})$ the open set defined in §5.2. Then

$\pi(U_4) \subset U_N$. Put $x = 2q_1q_2^{-1}$ and

$$\zeta_1 = z_1/z_4, \ \zeta_2 = z_2/z_4, \ \zeta_3 = z_3/z_4 ,$$

then

$$x = 2(\zeta_1 + j\zeta_2)(\zeta_3 + j)^{-1} ,$$

and hence

$$x^1 + x^2 i = \frac{2(\zeta_1\bar{\zeta}_3 + \bar{\zeta}_2)}{1 + |\zeta_3|^2} , \quad x^3 - x^4 i = \frac{2(\zeta_2\bar{\zeta}_3 - \bar{\zeta}_1)}{1 + |\zeta_3|^2} ,$$

so that

$$\begin{aligned} x^1 &= \frac{\zeta_1\bar{\zeta}_3 + \bar{\zeta}_1\zeta_3 + \zeta_2 + \bar{\zeta}_2}{1 + |\zeta_3|^2} , \quad x^2 = \frac{-i\,(\zeta_1\bar{\zeta}_3 - \bar{\zeta}_1\zeta_3 + \bar{\zeta}_2 - \zeta_2)}{1 + |\zeta_3|^2} , \\ x^3 &= \frac{\zeta_2\bar{\zeta}_3 + \bar{\zeta}_2\zeta_3 - \bar{\zeta}_1 - \zeta_1}{1 + |\zeta_3|^2} , \quad x^4 = \frac{i\,(\zeta_2\bar{\zeta}_3 - \bar{\zeta}_2\zeta_3 - \bar{\zeta}_1 + \zeta_1)}{1 + |\zeta_3|^2} . \end{aligned} \tag{5.3}$$

Let us define

$$2\, C^{\mu\nu}_{\bar{i}\bar{j}} = \frac{\partial x^\mu}{\partial \bar{\zeta}^i} \frac{\partial x^\nu}{\partial \bar{\zeta}^j} - \frac{\partial x^\nu}{\partial \bar{\zeta}^i} \frac{\partial x^\mu}{\partial \bar{\zeta}^j} .$$

Then, on U_4, $\pi^*\beta$ can be written

$$\pi^*\beta = \sum_{\mu\nu\bar{i}\bar{j}} \beta_{\mu\nu} \left\{ C^{\mu\nu}_{\bar{i}\bar{j}}\, d\bar{\zeta}^i \wedge d\bar{\zeta}^j + \frac{\partial x^\mu}{\partial \bar{\zeta}^i} \frac{\partial x^\nu}{\partial \zeta^j}\, d\bar{\zeta}^i \wedge d\zeta^j + \overline{C^{\mu\nu}_{\bar{i}\bar{j}}}\, d\zeta^i \wedge d\zeta^j \right\} . \tag{5.4}$$

It is straightforward to calculate that $C^{\mu\nu}_{\bar{i}\bar{j}}$ is anti-self-dual in its μ-ν indices, so that, if $\beta = {}^*\beta$, the first and last summations in (5.4) both vanish. This means that the pull-back of a self-dual 2-form to $P^3(\mathbb{C})$ involves only pairs of differentials of the type $d\bar{\zeta}^i \wedge d\zeta^j$. Such a form is said to be of type (1,1) and this property is independent of the coordinates used.

5.3. The notion of *type* for C^∞ differential forms on a complex manifold, which we met at the end of §5.2, is basic in the study of the differential geometry of these manifolds, and before proceeding with the construction of self-dual connections we shall investigate it further in a general setting. Thus let X be a complex n-dimensional manifold, and (U,ϕ) be a chart on X. Let

$$\phi(x) = (z^1(x), \ldots , z^n(x)) \in \mathbb{C}^n$$

define coordinate functions $z^1, \ldots , z^n$. Then

$$(\mathrm{Re}\, z^1, \mathrm{Im}\, z^1, \ldots , \mathrm{Re}\, z^n, \mathrm{Im}\, z^n)$$

define real C^∞ coordinates on U giving a 2n-dimensional C^∞ chart on X. It is easily checked that if we start with a complex n-dimensional atlas we are led to a 2n-dimensional oriented differentiable atlas.

Instead of using the real and imaginary parts of z^i, it is possible to use z^i and $\bar{z}^i$. The differentials

$$dz^1, d\bar{z}^1, \ldots, dz^n, d\bar{z}^n \tag{5.5}$$

are linearly independent on U (considered as a C^∞ manifold) and so any smooth form β may be expressed in terms of them. Any product of the differentials (5.5) can be written in the order

$$dz_{i_1} \wedge \ldots \wedge dz_{i_p} \wedge d\bar{z}_{j_1} \wedge \ldots \wedge d\bar{z}_{j_q} \tag{5.6}$$

with $i_1 < i_2 < \ldots < i_p$ and $j_1 < j_2 < \ldots < j_q$. A sum of products of differentials involving p unbarred and q barred differentials as in (5.6) is said to be of type (p,q). An m-form β can be written as a sum (in a unique fashion)

$$\beta = \beta_0 + \beta_1 + \ldots + \beta_m$$

with β_p of type (p, m-p) and this decomposition is independent of the (complex) coordinates used. The forms of type (p,q) form a vector space $\Omega^{p,q}$ and there is a direct sum decomposition

$$\Omega^m = \Omega^{0,m} + \Omega^{1,m-1} + \ldots + \Omega^{m,0} .$$

If f is a smooth function on X, df is a 1-form and therefore is a sum of forms of type (1,0) and (0,1). The part of df of type (1,0) is denoted by ∂f and the part of type (0,1) is denoted by $\bar{\partial} f$. The coefficient of $d\bar{z}^i$ in $\bar{\partial} f$ is written $\partial f/\partial \bar{z}^i$, and it is easily seen that

$$\bar{\partial} f = 0$$

expresses the Cauchy-Riemann equations in an invariant manner.

More generally, if β is in $\Omega^{p,q}$ then $d\beta$ is in $\Omega^{p+1,q} + \Omega^{p,q+1}$. The part of $d\beta$ in $\Omega^{p+1,q}$ is denoted by $\partial\beta$ and the part in $\Omega^{p,q+1}$ by $\bar{\partial}\beta$. This defines operators ∂ and $\bar{\partial}$ with properties similar to those of d:

$$\partial^2 = \bar{\partial}^2 = 0 , \qquad \partial\bar{\partial} + \bar{\partial}\partial = 0 ,$$

$$\bar{\partial}(\beta \wedge \gamma) = \bar{\partial}\beta \wedge \gamma + (-1)^m \beta \wedge \bar{\partial}\gamma,$$

if β is an m-form.

The operator $\bar{\partial}$ gives rise to complexes

$$\Omega^{m,0} \xrightarrow{\bar{\partial}} \Omega^{m,1} \xrightarrow{\bar{\partial}} \cdots \xrightarrow{\bar{\partial}} \Omega^{m,n}$$

called the Dolbeault complexes, and their cohomology groups are the *Dolbeault cohomology groups* of X. $\bar{\partial}$ satisfies the Dolbeault lemma: if $\bar{\partial}\beta = 0$ then on a neighbourhood of any given point $\beta = \bar{\partial}\gamma$ for some γ.

If E is a C^∞ vector bundle over the complex manifold X, it is said to be a holomorphic vector bundle if E is a complex manifold, and each point of X has a neighbourhood U such that there is a holomorphic map

$$\phi : \pi^{-1}(U) \to U \times \mathbb{C}^N$$

which is a diffeomorphism, and linear on each fibre. If E is a holomorphic vector bundle, a section s of E on an open set U is called holomorphic if it is a holomorphic map from U to E. The notions of holomorphic local frames and transition functions may be defined in the same manner as for C^∞ bundles.

If E is a C^∞ vector bundle over X and it has a collection of C^∞ local frames such that the transition functions are holomorphic, there is a unique holomorphic structure on E such that the given local frames are holomorphic.

If E is a C^∞ vector bundle over X, the bundle-valued forms on X have a decomposition into type, and a connection in E decomposes into a sum of two operators

$$D = D' + D''$$

with

$$D'' : \Omega^0(E) \to \Omega^{0,1}(E)$$

using an obvious notation. The curvature Ω of D splits into a sum of three terms

$$\Omega = \Omega_0 + \Omega_1 + \Omega_2$$

of types (2,0), (1,1) and (0,2) respectively. It is easy to check that

$$D''(D''s) = \Omega_2 \otimes s$$

for any s in $\Omega^0(E)$, so that Ω has no component of type (0,2) if and only if $(D'')^2 = 0$.

Now D" may be seen to satisfy the identity

$$D''(\beta \wedge \gamma) = (\bar{\partial}\beta) \wedge \gamma + (-1)^m \beta \wedge D''\gamma$$

for β in $\Omega^m(X)$ and γ in $\Omega^p(E)$. Let $(s_1, \ldots , s_N)$ be a local frame for E and define α_{ij} by

$$D''s_j = \sum_i \alpha_{ij} \otimes s_i ;$$

then α_{ij} is of type (0,1) and

$$D''(D''s_j) = \sum_i (\bar{\partial}\alpha_{ij} + \sum_k \alpha_{ik} \wedge \alpha_{kj}) \otimes s_i .$$

Thus Ω_2 vanishes provided

$$\bar{\partial}\alpha_{ij} + \sum_k \alpha_{ik} \wedge \alpha_{kj} = 0,$$

and this is the integrability condition for the equations

$$\bar{\partial}g_{ij} + \sum_k \alpha_{ik} g_{kj} = 0$$

to have a solution lying in $GL(N,\mathbb{C})$ in a neighbourhood of an arbitrary point; see Nijenhuis & Woolf (1963). If we define

$$t_j = \sum_i g_{ij} s_i$$

then

$$D''t_j = \sum_i (\bar{\partial}g_{ij} + \alpha_{ik} g_{kj}) s_i = 0,$$

and hence we see that the vanishing of Ω_2 is necessary and sufficient for the existence of local frames $(t_1, \ldots , t_N)$ satisfying

$$D''t_j = 0.$$

Hence $\Omega_2 = 0$ implies the existence of a unique holomorphic structure in E such that a local section s of E is holomorphic if and only if

$$D''s = 0.$$

Conversely, if E is holomorphic and in addition has a C^∞ Hermitian structure, there is a unique metric connection D such that $D''s = 0$ for any holomorphic section s. We call this the canonical connection of E.

From $\Omega_2 = 0$ it also follows that $(D'')^2 = 0$, so that

$$\Omega^0(E) \xrightarrow{D''} \Omega^{0,1}(E) \xrightarrow{D''} \Omega^{0,2}(E) \longrightarrow \dots \qquad (5.7)$$

is a complex. The cohomology groups $H^p(E)$ of this complex are very important invariants of the bundle E, and are called the Dolbeault cohomology groups of E.

On $P^3(\mathbb{C})$ it makes invariant sense to speak of algebraic vector bundles. These admit transition functions which are rational functions of the coordinates z_i. A theorem of Serre (1956) says that any holomorphic vector bundle on $P^3(\mathbb{C})$ is algebraic. (Serre's theorem is much more general than this.) This means that the methods of algebraic geometry may be used to study holomorphic vector bundles. The cohomology groups of an algebraic vector bundle defined by algebraic methods are isomorphic to the Dolbeault cohomology groups defined above when the bundle is regarded as a holomorphic vector bundle. See Griffiths & Adams (1974) for a nice treatment of the relationships between holomorphic and algebraic structures.

5.4. The following application of the results of the previous section is due to Atiyah & Ward (1977); see also Atiyah et al. (1977b).

If $\tilde{E}$ is a C^∞ Hermitian vector bundle over S^4 with a self-dual metric connection $\tilde{D}$ having self-dual curvature F, the pull-back $E = \pi^*\tilde{E}$ has the pulled-back connection $D = \pi^*\tilde{D}$ whose curvature $\Omega = \pi^*F$ is of type (1,1). Thus Ω has no component of type (0,2) and hence the (0,1) part D'' of D satisfies $(D'')^2 = 0$, and E has the structure of a holomorphic vector bundle whose holomorphic sections on an open set U are the solutions s of

$$D''s = 0.$$

Conversely, one may ask which Hermitian holomorphic vector bundles on $P^3(\mathbb{C})$ arise in this way? These are determined by two extra pieces of information. Firstly, if $E = \pi^*\tilde{E}$ as above, then D is flat when restricted to the fibres of π. The fibres are simply-connected, so E is trivialized on each fibre by covariant constant (and hence holomorphic) sections. Thus E is holomorphically trivial on each fibre.

Let $\sigma : P^3(\mathbb{C}) \to P^3(\mathbb{C})$ be defined as follows: If [z] is in $P^3(\mathbb{C})$ it determines (up to multiplication by complex scalars on the right) a pair (q_1, q_2) in $\mathbb{H}^2$. We let $\sigma[z]$ be determined by $(q_1 j, q_2 j)$. That is, if $\sigma[z] = [z']$, then

$$z_1' = -c\bar{z}_2,\ z_2' = c\bar{z}_1,\ z_3' = -c\bar{z}_4,\ z_4' = c\bar{z}_3$$

for some c in $\mathbb{C}$. [z] and $\sigma[z]$ lie in the same fibre of π, and the fibre of π through [z] is the line spanned by [z] and $\sigma[z]$. Since D is flat when restricted to a fibre, there is a map θ from $E_{[z]}$ to $E_{\sigma[z]}$ defined by parallel translation within the fibre. Define $\tau : E \to E^*$ (E^* is the dual bundle of E) by

$$\langle \tau v, w \rangle = (w, \theta v)_{\sigma[z]}$$

for v in $E_{[z]}$, w in $E_{\sigma[z]}$. τ is antilinear, and determines an isomorphism of $\sigma^*\bar{E}$ with E^*. This isomorphism τ of $\sigma^*\bar{E}$ with E^* is the second piece of information required; see Atiyah et al. (1977b) for details.

5.5. Hartshorne (1978) used the work of Atiyah & Ward (1977) and results from algebraic geometry to study the moduli spaces for small values of the charge. For k = 1 and 2 he showed the moduli spaces were connected and gave explicit geometric descriptions.

Later Atiyah, Hitchin, Drinfeld & Manin (1978) utilised the work of Horrocks (1964) and Barth & Hulek (1978) to give a simple algebraic construction of self-dual connections, and it is this construction which we shall describe in this final section.

There is a natural holomorphic line bundle H^* on $P^3(\mathbb{C})$ whose fibre at a point [z] is the 1-dimensional vector space spanned by z. The dual line-bundle H may be characterized, up to isomorphism, by $\dim H^0(H) = 4$. Let H^ℓ denote H tensored with itself ℓ times (if $\ell > 0$), or H^* tensored with itself $-\ell$ times (if $\ell < 0$). Then the vector space $\sum_{\ell \in \mathbb{Z}} H^1(E(\ell))$ as a module for the algebra $\sum_{\ell \in \mathbb{Z}} H^0(H^\ell)$ determines E up to isomorphism, where $E(\ell)$ denotes $E \otimes H^\ell$. Barth & Hulek showed that if $H^1(E(-2)) = 0$, and if E restricted to some line in $P^3(\mathbb{C})$ is holomorphically trivial, then this module is especially simple and is determined by the action on $H^1(E(-1))$.

The main observation of Atiyah et al. (1978) is that $H^1(E(-2)) = 0$ for instanton bundles, and E is trivial on the fibres of π, which are lines, so that the instanton bundles are described by the Barth & Hulek result. A proof that $H^1(E(-2)) = 0$ using the complex (5.7) may be found in Rawnsley (1978a). The method of this proof is to show that if $D''\alpha = 0$ with α in $\Omega^{0,1}(E(-2))$ then $\alpha = \psi(s)$ where $\psi : \Omega^0(\tilde{E}) \to \Omega^{0,1}(E(-2))$ is an explicitly constructed map. Further

$$D''\psi(s) = \pi^*((D^*D + R/6)s) \otimes \beta$$

where β is a certain form in $\Omega^{0,2}(H^{-2})$. Then $H^1(E(-2))$ becomes identical with the kernel of $D^*D + R/6$ on $\tilde{E}$. Since D^*D is non-negative and the scalar curvature R of the standard metric on S^4 is positive, it follows that $D^*D + R/6$ is positive-definite, and so has no kernel, and hence $H^1(E(-2)) = 0$.

For simplicity we shall give the construction for rank two vector bundles, which correspond with $SU(2)$ instantons. Let W denote the dual of $H^1(E(-1))$ and V denote $H^1(E \otimes \Omega^1)$ where Ω^1 denotes the holomorphic cotangent bundle of $P^3(\mathbb{C})$. Then $\dim W = k$ and $\dim V = 2k + 2$. Moreover $E \cong E^*$ (because $c_1(E) \cong c_1(\tilde{E}) = 0$ implies $\Lambda^2 E$ is trivial, and then exterior multiplication $E \times E \to \Lambda^2 E$ sets up the isomorphism), and this gives a bilinear form $\langle \cdot , \cdot \rangle$ on V which is symplectic. The isomorphism $\sigma^*\bar{E} \cong E^*$ gives a conjugation τ on V such that $\tau^2 = -1$ and

$$\langle v, \tau w \rangle \quad , \quad v, w \in V$$

defines a positive-definite inner product on V. Choosing a basis for V which is simultaneously symplectic and orthonormal, we may assume $V = \mathbb{C}^{2k+2}$, and

$$\langle v, w \rangle = -v^T J w, \quad \tau w = J\bar{w} \quad , \quad v, w \in \mathbb{C}^{2k+2} ,$$

where

$$J = \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix}$$

in terms of $(k+1) \times (k+1)$ blocks, and v^T denotes the transpose of the column v.

The module structure is described by a family

$$A(z) = \sum_{i=1}^{4} z_i A_i \quad , \qquad z \in \mathbb{C}^4 ,$$

of $k \times (2k+2)$ matrices which must satisfy

$$\begin{array}{lll} \text{(i)} \quad A(z)^T J\, A(z) = 0 , & \text{for all } z; & \\ \text{(ii)} \quad J\, \overline{A(z)} = A(jz) , & \text{for all } z; & (5.8) \\ \text{(iii)} \quad \text{rank } A(z) = k , & \text{for all } z \neq 0. & \end{array}$$

Here

$$j \cdot z = \begin{pmatrix} -\bar{z}_2 \\ \bar{z}_1 \\ -\bar{z}_4 \\ \bar{z}_3 \end{pmatrix} \qquad \text{if} \qquad z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} .$$

If we write

$$A_i = \begin{pmatrix} B_i \\ C_i \end{pmatrix} \quad , \qquad i = 1, 2, 3, 4,$$

where B_i, C_i are $k \times (k+1)$ matrices, then (5.8 ii) implies $A_2 = J\bar{A}_1$, $A_4 = J\bar{A}_3$, or

$$B_2 = \bar{C}_1, \quad C_2 = -\bar{B}_1, \quad B_4 = \bar{C}_3, \quad C_4 = -\bar{B}_3 ;$$

Let us identify $\mathbb{C}^{2k+2}$ with $\mathbb{H}^{k+1}$ by means of the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \longrightarrow x + jy \qquad x, y \in \mathbb{C}^{k+1} ,$$

and similarly identify $k \times (2k+2)$ complex matrices with $k \times (k+1)$ quaternion matrices; then $A(z)$ becomes

$$\begin{aligned} & z_1 B_1 + z_2 B_2 + z_3 B_3 + z_4 B_4 + j(z_1 C_1 + z_2 C_2 + z_3 C_3 + z_4 C_4), \\ = \; & (B_1 + jC_1)(z_1 + jz_2) + (B_3 + jC_3)(z_3 + jz_4), \\ = \; & R_1 q_1 + R_2 q_2 , \end{aligned}$$

with

$$q_1 = z_1 + jz_2, \; q_2 = z_3 + jz_4 , \quad R_1 = B_1 + jC_1 , \quad R_2 = B_3 + jC_3 .$$

Thus the quaternionic family $R(q) = R_1 q_1 + R_2 q_2$, $q \in \mathbb{H}^2$, determines the bundle E. Moreover the conditions (5.8) may be combined into the quaternionic form

$$R(q)^* R(q) \in GL(k, \mathbb{R}), \qquad \text{for all } q \in \mathbb{H}^2, \; q \neq 0. \qquad (5.9)$$

The theory of Barth & Hulek also indicates, in terms of the family $A(z)$, when two bundles are isomorphic. Namely, $A(z)$ and $A'(z)$ come from isomorphic bundles if and only if there is T in $GL(k, \mathbb{R})$ and S in $Sp(2k+2)$ such that

$$A'(z) = S\,A(z)\,T\ , \quad \text{for all } z \in \mathbb{C}^4. \tag{5.10}$$

Thus the moduli space may be described as the set of families $A(z)$ modulo the equivalence relation determined by (5.10). Identifying $Sp(2k+2)$ as the (quaternionic) unitary group of $\mathbb{H}^{k+1}$, this moduli space is the same as the space of families $R(q)$ satisfying (5.9), where two families $R(q)$, $R'(q)$ are equivalent if

$$R'(q) = S\ R(q)\ T \tag{5.11}$$

with T in $GL(k,\mathbb{R})$, S in $Sp(2k+2)$.

The bundle E may be constructed from $A(z)$ (see Atiyah et al. (1978)), and then $\tilde{E}$ and $\tilde{D}$ are obtained by reversing the construction of Atiyah & Ward. Instead of this it is possible to construct the $SU(2)$ connection from $R(q)$ directly, utilising the isomorphism $S^4 \cong P^1(\mathbb{H})$. If $[q_1,q_2]$ is in S^4, let

$$P_{[q_1,q_2]} = \{v \in \mathbb{H}^{k+1} \mid R(q)^* v = 0,\ v^*v = 1;\ (q_1,q_2) = q\}\ .$$

Then P is a principal $SU(2)$ bundle where we identify $SU(2)$ with the group of unit quaternions ($\{u \in \mathbb{H} \mid u^*u = 1\}$), and $SU(2)$ acts on P by

$$v \cdot u = vu\ , \qquad v \in P,\ u \in SU(2).$$

P is clearly a subset of $S^4 \times \{v \in \mathbb{H}^{k+1} \mid v^*v = 1\}$, and inherits a connection from the trivial connection in this product bundle.

Within the equivalence class determined by (5.11) there is a subset with R_1 of the form

$$R_1 = \begin{array}{c} k \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{array}\begin{array}{c} \\ k \\ 1 \end{array} .$$

Let

$$R_2 = \begin{pmatrix} X \\ y \end{pmatrix}$$

where X is $k \times k$ and y is $1 \times k$. Put $x = q_1 q_2^{-1}$ (x is defined on U_N) then, for $[q_1,q_2] = [x,1]$ in U_N ,

$$P_{[q_1,q_2]} = \{ \begin{pmatrix} v \\ u \end{pmatrix} \in \mathbb{H}^{k+1} \mid (x + X)^* v + y^* w = 0, \quad v \in \mathbb{H}^k,\ w \in \mathbb{H}\}\ .$$

If x is a point of U_N with $(x + X)$ invertible, it follows we have a local section s of P given by

$$s([q_1,q_2]) = \begin{pmatrix} -(x + X)^{-1*}\ y^*c \\ c \end{pmatrix}, \quad c = (1 + y(x + X)^{-1}(x + X)^{-1*}\ y^*)^{-\frac{1}{2}}\ ;$$

then

$$A_s = s^*ds$$

gives the gauge potential where s is defined. Put

$$v(x) = -(x + X)^{-1*}\ y^*$$

so that

$$c^{-2} = 1 + v^*v,$$

then

$$A_s = \frac{v^*dv - dv^*v}{2(1 + v^*v)}$$

which is a rational function of x.

Further details of this quaternionic description of instantons may be found in Rawnsley (1978b), Corrigan, Fairlie, Templeton & Goddard (1978), and Christ, Weinberg & Stanton (1978).

Appendix: Characteristic Classes and Chern-Weil Theory

A.1. Let G be a reductive Lie group and $\mathfrak{g}$ its Lie algebra with a non-singular *invariant bilinear form* $(\cdot,\cdot)$. A function $Q : \mathfrak{g} \to \mathbb{R}$ is a *polynomial* if, with respect to any basis of $\mathfrak{g}$, Q is a polynomial function of the coordinates. Such a polynomial is *invariant* if

$$Q(g\, X\, g^{-1}) = Q(X) \qquad \text{for all } X \in \mathfrak{g},\ g \in G. \tag{A.1}$$

Let $I(\mathfrak{g})$ denote the ring of invariant polynomials, and $I^k(\mathfrak{g})$ the subset of elements of degree k, k = 0, 1, ...

Given a smooth function $Q : \mathfrak{g} \to \mathbb{R}$, we define Q' by

$$(Q'(X), Y) = \frac{d}{dt} Q(X + t\, Y)\Big|_{t=0} .$$

From (A.1) it follows at once that if Q is in $I(\mathfrak{g})$

$$Q'(g\, X\, g^{-1}) = Q'(X) g^{-1} . \tag{A.2}$$

There is also a chain rule: If X depends on a parameter s, say, then

$$\frac{d}{ds} Q(X(s)) = (Q'(X(s)), X'(s)). \tag{A.3}$$

Differentiating again we may define $Q''(X)$ as a linear map from $\mathfrak{g}$ to itself by

$$Q''(X)\, Y = \frac{d}{dt} Q'(X + t\, Y)\Big|_{t=0} . \tag{A.4}$$

Then

$$\frac{d}{ds} Q'(X(s)) = Q''(X(s))\, X'(s) . \tag{A.5}$$

Applying (A.5) to $X(s) = \exp s\,Y\, X \exp -s\, Y$ when Q is in $I(\mathfrak{g})$, and setting s equal to 0, we get

$$Q''(X)([Y,X]) = [Y, Q'(X)]. \tag{A.6}$$

Putting X = Y gives

$$[X, Q'(X)] = 0. \tag{A.7}$$

Clearly these results are still true if we tensor $\mathfrak{g}$ with any commutative $\mathbb{R}$-algebra, and in particular we may take the algebra $\sum_{p\geq 0} \Omega^{2p}(X)$ of forms of even degree on a manifold. If $F \in \Omega^2(X) \otimes \mathfrak{g}$ and $Q \in I^k(\mathfrak{g})$, then $Q(F) \in \Omega^{2k}(X)$ and the chain rule (A.3) becomes

$$d\, Q(F) = (Q'(F), dF). \tag{A.8}$$

Also $Q'(F) \in \Omega^{2(k-1)}(X) \otimes \mathfrak{g}$.

A.2. Let (P,π,G) be a principal G-bundle over a manifold X, and A a connection in P. For each section $s : U \to P$, U open in X, we have $F_s \in \Omega^2(U) \otimes \mathfrak{g}$, and hence each $Q \in I^k(\mathfrak{g})$ determines $Q(F_s) \in \Omega^{2k}(U)$. Then the invariance of Q implies

$$Q(F_s) = Q(F_t)$$

if $t = s\cdot g$ with $g : U \to G$. Thus $Q(F_s)$ is the restriction to U of a globally, invariently defined 2k-form which we denote by Q_A. (A.8) implies

$$d\, Q_A|U = (Q'(F_s), d\, F_s) .$$

By Bianchi's identity $d\, F_s = -\langle A_s, F_s\rangle$, and by the invariance of the inner product, we obtain

$$d\, Q_A|U = ([Q'(F_s),F_s], A_s),$$

which vanishes by (A.7). Thus Q_A is a closed form, and so defines a cohomology class in $H^{2k}(X)$.

Suppose we have two connections A, A' in P, and $A' = A + a$ with a in $\Omega^1(X,P(\mathfrak{g}))$. Define

$$A^t = A + t\, a,$$

so A^t is a curve of connections joining A to A', and let A^t have curvature F^t. As in §4,

$$F^t = F + t\, Da + \frac{t^2}{2}\langle a,a\rangle ,$$

where F is the curvature of A. Hence

$$\frac{d\,F^t}{dt} = Da + t\langle a,a\rangle = D^t\, a$$

where D^t is the covariant exterior derivative defined by A^t.

Then for any Q in $I^k(\mathfrak{g})$,

$$\begin{aligned}(Q_{A'} - Q_A)|U &= \int_0^1 \frac{d}{dt} Q_{A^t}\, dt|U \\ &= \int_0^1 \frac{d}{dt} Q(F_s^t)\, dt \\ &= \int_0^1 (Q'(F_s^t), D^t\, a_s)\, dt.\end{aligned}$$

On the other hand

$$d(Q'(F_s^t), a_s) = (D^t Q'(F_s^t), a_s) + (Q'(F_s^t), D^t a_s)$$

$$= (Q''(F_s^t) D_s^t F_s^t, a_s) + (Q'(F_s^t), D^t a_s),$$

and the first term on the right-hand side is zero from Bianchi's identity. Thus

$$(Q_{A'} - Q_A)|U = d \int_0^1 (Q'(F_s^t), a_s)\, dt.$$

Again, $(Q'(F_s^t), a_s)$ is independent of s, and so is globally defined; and hence we have obtained a form $Q_{A,A'}$ with

$$Q_{A'} - Q_A = d\, Q_{A,A'} .$$

Hence the class $[Q_A]$ is independent of our choice of connection in (P,π,G). We thus have a map

$$q_P : I(\mathfrak{g}) \to H^*(X)$$

associated to (P,π,G) given by

$$q_P(Q) = [Q_A]$$

for any connection A in P. q_P is the *Chern-Weil homomorphism*, and its image is the *characteristic ring* of (P,π,G). The classes $q_P(Q)$ are called *characteristic classes*, and their values on cycles in X are *characteristic numbers*.

A.3. Let G be a group of n x n matrices and define $c_k \in I^k(\mathfrak{g})$ by

$$\mathrm{Det}\left[\lambda I + \frac{X}{2\pi i}\right] = \sum_{k=0}^{n} \lambda^{n-k} c_k(X) .$$

We put

$$c_k(P) = q_P(c_k) , \quad k = 1, 2, \dots ,$$

and call $c_k(P)$ the *k-th Chern class* of P. For G = SU(2), $c_1(X) = \mathrm{Tr}[X/2\pi i] = 0$, by definition, so $c_1(P) = 0$, whilst $c_2(X) = \frac{1}{8\pi^2}\mathrm{Tr}[X^2]$. On S^4 we then obtain the second Chern number

$$c_2(P) = \frac{1}{8\pi^2} \int_{S^4} \mathrm{Tr}\,[F_s \wedge F_s]$$

$$= \frac{1}{\cancel{8}\pi^2} \int_{S^4} \sum_{\mu\nu} \mathrm{Tr}\,[F_{\mu\nu}\tilde{F}^{\mu\nu}] ,$$

16

where

$$\tilde{F}^{\mu\nu} = \tfrac{1}{2}\sum_{\rho\sigma} \varepsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} ,$$

and

$$F_s = \tfrac{1}{2} \sum_{\mu\nu} F_{\mu\nu}\, dx^\mu \wedge dx^\nu .$$

If, instead, we define

$$ch(X) = \mathrm{Tr}\,[\exp X/2\pi i],$$

we obtain $ch(P) \in H^*(X)$ (which is actually a polynomial when applied to a form, since $\Omega^p(X) = 0$ for $p > \dim X$). ch(P) is the Chern character of P; $ch_k(P)$ is the component in degree 2k and is given by

$$ch_k(X) = \frac{1}{k!} \mathrm{Tr}\,[\left(\frac{X}{2\pi i}\right)^k] ;$$

and in particular

$$ch_0(X) = n.$$

For further details see Greub, Halperin & Vanstone (1972), and Dupont (1978).

REFERENCES

Ambrose, W. & Singer, I. M. 1953 *Trans. Amer. Math. Soc.* 75, 428 - 443.

Atiyah, M. F., Bott, R., Patodi, V. K. 1973 *Invent. Math.* 19, 279 - 330.

Atiyah, M. F., Hitchin, N. J., Drinfeld, V. G., Manin, Yu. I. 1978 *Phys. Letters A* 65, 185 - 187.

Atiyah, M. F., Hitchin, N. J., Singer, I. M. 1977a *Proc. Nat. Acad. Sci. U.S.A.* 74, 2662.

——— 1977b *Self-duality in Riemannian geometry.* Oxford preprint: To appear in the Proc. Roy. Soc.

Atiyah, M. F. & Ward, R. S. 1977 *Commun. Math. Phys.* 55, 117 - 124.

Barth, W. & Hulek, K. 1978 *Monads and moduli of vector bundles.* Erlangen preprint.

Belavin, A. A., Polyakov, A. M., Schwartz, A. S., Tyupkin, S. 1976 *Phys. Letters B* 59, 85 - 87.

Bernard, C. W., Christ, N. H., Guth, A. H., Weinberg, E. J. 1977 *Instanton parameters for arbitrary gauge groups.* Columbia preprint CU-TP-105.

Chern, S. S. 1967 *Complex manifolds without potential theory.* Princeton, Van Nostrand.

Christ, N. H., Weinberg, E. J., Stanton, N. K. 1978 *General self-dual Yang-Mills solutions.* Columbia preprint CU-TP-119.

Coleman, S. 1978 *The uses of instantons.* Harvard preprint HUTP-78/1004.

Corrigan, E. F., Fairlie, D. B., Templeton, S., Goddard, P. 1978 *A Green's function for the general self-dual gauge field.* Paris preprint LPTENS 78/13.

Dupont, J. L. 1978 *Curvature and characteristic classes.* Berlin, Springer-Verlag. (Lecture Notes in Mathematics 640).

Greub, W., Halperin, S., Vanstone, R. 1972-1975 *Connections, curvature and cohomology.* 3 vols. New York, Academic Press.

Griffiths, P. & Adams, J. 1974 *Topics in algebraic and analytic geometry.* Princeton, Princeton University Press. (Mathematical Notes 13).

Hartshorne, R. 1977 *Algebraic Geometry.* New York, Springer-Verlag.

——— 1978 *Commun. Math. Phys.* 59, 1 - 15.

Horrocks, G. 1964 *Proc. London Math. Soc.* 14, 689 - 713.

Husemoller, D. 1975 *Fibre bundles.* 2nd edition. New York, Springer-Verlag.

Jackiw, R., Nohl, C., Rebbi, C. 1977 *Phys. Rev. D* 15, 1642 - 1654.

Jackiw, R., Rebbi, C. 1977 *Phys. Letters B* 67, 189 - 192.

Kobayashi, S. & Nomizu, K. 1963 *Foundations of differential geometry.* vol. I, New York, Wiley.

Kuranishi, M. 1965 *New proof for the existence of locally complete families of complex structures.* Proceedings of the Conference on Complex Analysis, Minneapolis 1964. New York, Springer-Verlag.

Loos, H. 1967 *J. Mathematical Phys.* 8, 2114 - 2124.

Morrow, J. & Kodaira, K. 1971 *Complex manifolds.* New York, Holt, Reinhart & Winston.

Nijenhuis, A. & Woolf, W. 1963 *Ann. of Math.* 77, 424 - 489.

Penrose, R. 1977 *Reports on Mathematical Phys.* 12, 65 - 76.

Rawnsley, J. H. 1978a *On the Atiyah-Hitchin-Drinfeld-Manin vanishing theorem for cohomology groups of instanton bundles.* Dublin Institute preprint DIAS-TP-78-32.

——— 1978b *Self-dual Yang-Mills fields.* Dublin Institute preprint DIAS-TP-78-38. To appear in the Proceedings of the Seminar on Global Analysis, Canadian Mathematical Congress, Calgary 1978.

Schwarz, A. S. 1977 *Phys. Letters B* 67, 172 - 174.

Serre, J.- P. 1956 *Ann. Inst. Fourier* (Grenoble) 6, 1 - 42.

Wells, R. O. 1977 *Complex manifolds and mathematical physics.* Rice University preprint.

Yang, C. N. & Mills, R. 1954 *Phys. Rev.* 96, 191 - 195.

Yang, C. N. & Wu, T. T. 1975 *Phys. Rev. D* 12, 3845 - 3857.